AMC

AMCP 706-136

THIS IS A REPRINT WITHOUT CHANGE OF ORDP 20-136, REDESIGNATED AMCP 706-136

4-" 1Y/6

ENGINEERING DESIGN HANDBOOK

SERVOMECHANISMS

SECTION 1, THEORY

HEADQUARTERS, U. S. ARMY MATERIEL COMMAND

APRIL 1965

HEADQUARTERS
UNITED STATES ARMY MATERIEL COMMAND
WASHINGTON, D. C. 20315

30 April 1965

AMCP 706-136, Servomechanisms, Section 1, Theory, forming part of the Army Materiel Command Engineering Design Handbook Series, is published for the information and guidance of all concerned.

(AMCRD)

FOR THE COMMANDER:

SELWYN D. SMITH, JR.
Major General, USA
Chief of Staff

OFFICIAL:

STANLEY J. SAWICKI
Colonel, GS
Chief, Administrative Office

DISTRIBUTION: Special

FOREWORD

INTRODUCTION

This is one of a group of handbooks covering the engineering information and quantitative data needed in the design and construction of military equipment, which (as a group) constitutes the Army Materiel Command Engineering Design Handbook Series.

PURPOSE OF HANDBOOK

The handbook on Servomechanisms has been prepared as an aid to designers of automatic control systems for Army equipments, and as a guide to military and civilian personnel who are responsible for setting control-system specifications and ensuring their fulfillment.

SCOPE AND USE OF HANDBOOK

The publications are presented in handbook form rather than in the style of textbooks. Tables, charts, equations, and bibliographical references are used in abundance. Proofs and derivations are often omitted and only final results with interpretations are stated. Certain specific information that is always needed in carrying out design details has, of necessity, been omitted. Manufacturers' names, product serial numbers, technical specifications, and prices are subject to great variation and are more appropriately found in trade catalogs. It is essential that up-to-date catalogs be used by designers as supplements to this handbook.

To make effective use of the handbook during the design of a servo, the following procedure is suggested. The designer should turn first to Chapters **16** and **17** where design philosophy and methods are discussed. Implementation of the design procedure may require a review of certain theoretical concepts and methods which can be achieved through reference to Chapters **1** through **10.** As the design proceeds, a stage will be reached at which the power capacity of the output member has been fixed. Reference to Chapters **14, 15,** and **16** will then illustrate the salient features of output members having the required power capacity. After the designer has chosen the output member, he will find the information dealing with sensing elements and amplifiers (Chapters **11, 12,** and **13)** helpful in completing the design.

FEEDBACK CONTROL SYSTEMS AND SERVOMECHANISMS

Servomechanisms are part of a broad class of systems that operate on the principle of feedback. In a feedback control system, the output (response) signal is made to conform with the input (command) signal by feeding back to the input a signal that is a function of the output for the purpose of comparison. Should an error exist, a corrective action is automatically initiated to reduce the error toward zero. Thus, through feedback, output and input signals are made to conform essentially with each other.

In practice, the output signal of a feedback control system may be an electrical quantity such as a voltage or current, or any one of a variety of physical quantities such as a linear or angular displacement, velocity, pressure, or temperature. Similarly, the input signal may take any one of these forms. Moreover, in many applications, input signals belong to one of these types, and the output to another. Suitable transducers or measuring devices must then be used. It is also common to find multiple feedback paths or loops in complicated feedback control systems. In these systems, the over-all system performance as characterized by stability, speed of response, or accuracy can be enhanced by feeding back signals from various points within the system to other points for comparison and initiation of correction signals at the comparison points.

At present, there is no standard definition of a servomechanism. Some engineers prefer to classify any system with a feedback loop as a servomechanism. According to this inter-

pretation, an electronic amplifier with negative feedback is a servo. More frequently, however, the term servgmechanism is reserved for a feedback control system containing a mechanical quantity. Thus, the IRE defines a servomechanism as "a feedback control system in which one or more of the system signals represents mechanical motion." Some would restrict the definition further by applying the term only to a special class of feedback control system in which the output is a mechanical position.

APPLICATION OF SERVOMECHANISMS TO ARMY EQUIPMENT

Servomechanisms are an important part of nearly every piece of modern mechanized Army equipment. They are used to automatically position gun mounts, missile launchers, and radar antennas. They aid in the control of the flight paths of jet-propelled rockets and ballistic missiles, and play an important role in the navigational systems of those vehicles. As instrument servos, they permit remote monitoring of physical and electrical quantities and facilitate mathematical operations in computers.

No single set of electrical and physical requirements can be stated for servomechanisms intended for these diverse military applications. The characteristics of each servomechanism are determined by the function it is to perform, by the characteristics of the other devices and equipments with which it is associated, and by the environment to which it is subjected. It will often be found that two or more servo-system configurations will meet a given set of performance specifications. Final choice of a system may then be determined by such factors as ability of the system to meet environmental specifications, availability of components, simplicity, reliability, ease of maintenance, ease of manufacture, and cost. Finally, the ability to translate any acceptable paper design into a piece of physical equipment that meets electrical and physical specifications and works reliably depends to a great extent upon the skill of the engineering and manufacturing groups responsible for building the system. The exercise of care and good judgment when specifying electrical, mechanical, and thermal tolerances on components and subsystems can contribute greatly to the successful implementation of servo-system design.

The handbook on Servomechanisms was prepared under the direction of the Engineering Handbook Office, Duke University, under contract to the Army Research Office-Durham. The material for this pamphlet was prepared by Jackson & Moreland, Boston, Massachusetts, under subcontract to the Engineering Handbook Office. Jackson & Moreland was assisted in their work by consultants who are recognized authorities in the field of servomechanisms.

PREFACE

The Engineering Design Handbook Series of the Army Materiel Command is a coordinated series of handbooks containing basic information and fundamental data useful in the design and development of Army materiel and systems. The handbooks are authoritative reference books of practical information and quantitative facts helpful in the design and development of Army materiel so that it will meet the tactical and the technical needs of the Armed Forces. The present handbook is one of a series on Servomechanisms.

Section 1 of the handbook contains Chapters 1 through 10, which present feedback control theory as related to servomechanisms. This material is a concise summary of information on the subject. For this reason, persons who are unfamiliar with servomechanisms theory may find it necessary at first to acquaint themselves with the material included in standard textbooks. The bibliography at the end of each chapter lists applicable textbooks and periodicals for additional referencing and research.

For information on servomechanism components and system design, see one of the following applicable sections of this handbook:

AMCP 706-137	Section 2 Measurement and Signal Converters (Chapters 11-12)
AMCP 706-138	Section 3 Amplification (Chapter 13)
AMCP 706-139	Section 4 Power Elements and System Design (Chapters 14-20)

An index for the material in all four sections is placed at the end of Section 4.

Elements of the U. S. Army Materiel Command having need for handbooks may submit requisitions or official requests directly to Publications and Reproduction Agency, Letterkenny Army Depot, Chambersburg, Pennsylvania 17201. Contractors should submit such requisitions or requests to their contracting officers.

Comments and suggestions on this handbook are welcome and should be addressed to Army Research Office-Durham, Box CM, Duke Station, Durham, North Carolina 27706.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont)

TABLE OF CONTENTS (cont)

TABLE OF CONTENTS (cont)

TABLE OF CONTENTS (cont)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (cont)

LIST OF ILLUSTRATIONS (cont)

LIST OF ILLUSTRATIONS (cont)

LIST OF ILLUSTRATIONS (cont)

LIST OF ILLUSTRATIONS (cont)

LIST OF ILLUSTRATIONS (cont)

LIST OF TABLES

CHAPTER 1

PROPERTIES OF FEEDBACK CONTROL SYSTEMS*

1-1 OBJECTIVES OF A FEEDBACK CONTROL SYSTEM

The purpose of a feedback control system is to monitor an output (controlled variable) in a manner dictated by an input (reference variable) in the presence of spurious disturbances (such as random load changes). The basic elements of a feedback control system are shown in Fig. **1-1.** The system measures the output, compares the measurement with the desired value of the output as prescribed by the input, and uses the error (difference between actual output and desired output) to change the actual output and bring it into closer correspondence with the desired value of the output. To achieve a more sensitive control means, the error is usually amplified ; in general, the higher the gain the more accurate the system. Thus, a feedback control system is characterized by measurement, comparison, and amplification. In brief, a feedback control system is an error-correcting power-amplifying system that produces a high-accuracy output in accordance with **the** dictates of a prescribed input.

Since arbitrary disturbances (such as amplifier drift, random torques, etc) can occur at various points in the system, a feedback control system must be able to perform its task with the required accuracy in the presence of these disturbances. Since random noise (unwanted fluctuations) often is present at the input of the system, a feedback control system must be able to reject, or filter out, the noise while producing as faithful a representation of the desired output as is feasible.

By L. A. Gould

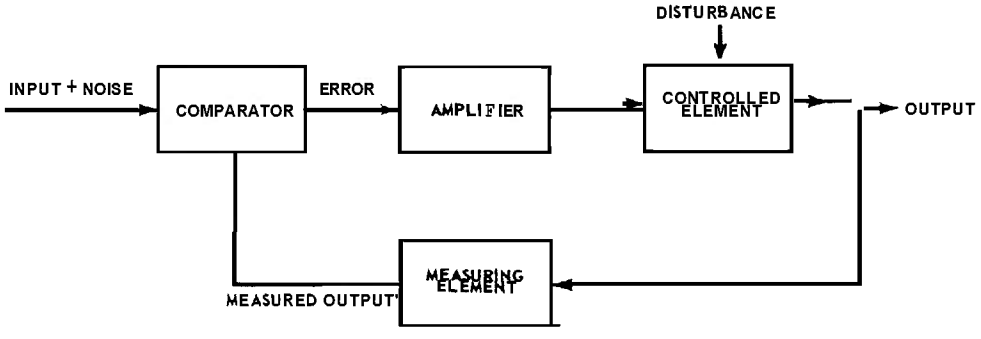


Fig. 1-1 Elements of a feedback control system.

1-2 OPEN-LOOP VS CLOSED-LOOP SYSTEM CHARACTERISTICS

Because a measure of the output is fed back and compared with the input, any representation of a feedback system contains a closed loop (see Fig. 1-1), and the system is thus called a *closed-loop* system. Many control systems do not exhibit this closed-loop feature and may be termed *open-loop* systems. In an open-loop system, the error is reduced by careful calibration. The elements of an open-loop control system are shown in Fig. 1-2.

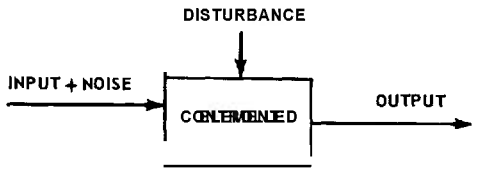


fig. 7-2 Elements of an open-loop control system.

If open-loop and closed-loop systems are compared, it can be seen that several advantages accrue to the closed-loop system. In a closed-loop system, the percentage change in the response of the system to a given percentage change in the response of one of its elements is approximately inversely proportional to the over-all amplification of the loop. However, in an open-loop system, the percentage change in the response of the system is approximately proportional to the percentage change in the response of one of its elements. Thus, a feedback control system is insensitive to changes in the parameters of its components and can usually be constructed from less accurate and cheaper components than those used in an open-loop system. One exception to the foregoing statement results from an inherent limitation — the closed-loop system can be no more accurate and reliable than its measuring element. The same limitation holds true for an open-loop system.

The error produced in an open-loop system by a given disturbance is much larger than the error produced by the same disturbance in an equivalent closed-loop system, the ratio of errors being approximately proportional to the over-all amplification of the loop of the closed-loop system. Thus, a feedback control system is relatively insensitive to extraneous disturbances and can be used in situations where severe upsets are expected. One can conclude that the over-all amplification (or gain) that can be achieved inside the loop of a feedback control system directly affects the accuracy of the system, the constancy of its characteristics, and the "stiffness" of the system in the face of external upsets or disturbances. In general, it is found that the higher the gain of the system, the better the system. The highest gain that can be used, however, is limited in every case by considerations of stability.

1-3 STABILITY AND DYNAMIC RESPONSE

For the advantages of accuracy and constancy of characteristics, the feedback control system must pay a price in the form of a greater tendency toward instability. A linear system is said to be stable if the response of the system to any discontinuous input does not exhibit sustained or growing oscillations. Essentially, this means that the system response will ultimately settle down to some steady value. An unstable system which exhibits steady or runaway oscillations is unacceptable. Unstable behavior must be guarded against in the design, construction, and testing of feedback systems. Because of the possibility of instability, a major portion of control system design is devoted to the task of ensuring that a safe margin of stability exists and can be maintained throughout the operating range of the system.

It can be shown that the cause of instability in a given closed-loop system is due to the fact that no physical device can respond instantaneously to a sudden change at its input. If a sudden change occurs in the error of a

feedback system, the output will not correct for the error instantaneously. If the corrective force is great enough (due to a high amplification), the output will accelerate rapidly and cause a reversal of the error. If a high output velocity is attained, the inertia of the output will carry the output past the point where the error is zero. Instability occurs if the maximum magnitude of the error after reversal is equal to, or greater than, the magnitude of the original disturbance in the error. The tendency for a system to become unstable is accentuated as the amplification is increased, since the stored energy in the inertia of the output will be correspondingly increased without any compensating increase in the rate of dissipation of energy in the system. This situatiop corresponds to an excessive delay in the response of the output. Thus, an attempt to increase accuracy by increasing gain or amplification is usually accompanied by an increased tendency toward instability. As a result, design becomes a compromise between accuracy and stability. A more detailed and quantitative examination of stability is developed in Ch. **4.**

The dynamic behavior of a system is important not only as a determinant of stability but also, for stable systems, as a measure of instantaneous accuracy. In many situations where rapid input variations occur, it is of the utmost importance that the error be kept within specified bounds ***at all times.*** Ideally, a system with no time lag would be able to follow extremely rapid input variations with perfect accuracy at all times. Actually, the impossibility of achieving instantaneous response, together with the stability problem created by the "pile-up" of the dynamic lags of cascaded elements in a loop, make the problem of maintaining dynamic accuracy (i.e., error within specified bounds ***at all times)*** progressively more difficult as the rapidity of input variations increases. Consequently, the design of both system and components is focused to a large degree on improving the speed of response (in other words, reducing dynamic lags), thereby obtaining a concomitant improvement in the over-all dynamic accuracy of the system.

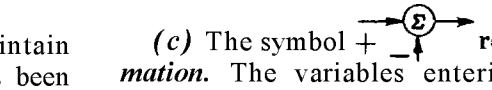
1-4 TERMINOLOGY OF FEEDBACK CONTROL SYSTEMS

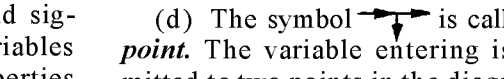
To facilitate discussion and to maintain uniformity, a specific terminology has been adopted. The general diagram of a feedback control system is shown in Fig. **1-3.** Note that some of the elements and variables in this diagram correspond to real devices and signals, whereas other elements and variables correspond to purely hypothetical properties of the system that are useful in visualizing the various functions of the system.

To aid visualization and to distinguish between variables and components, the symbolism of Fig. **1-3** is defined as follows :

(a) **A** line represents a ***variable*** or ***signal.*** The arrow on the line designates the direction **of** information flow.

(b) **A** block represents a ***device*** **or** ***group of devices*** that operate on the signal or signals entering the block to produce the signal leaving the block.

(c) The symbol  **represents** ***summation.*** The variables entering are added algebraically, according to the signs associated with the corresponding arrows, to produce the variable leaving.

(d) The symbol  is called a ***splitting point.*** The variable entering is to be transmitted to two points in the diagram. The variables leaving are both ***identical*** to the variable entering.

The symbolism used for the variables in Fig. **1-3** is defined as follows:

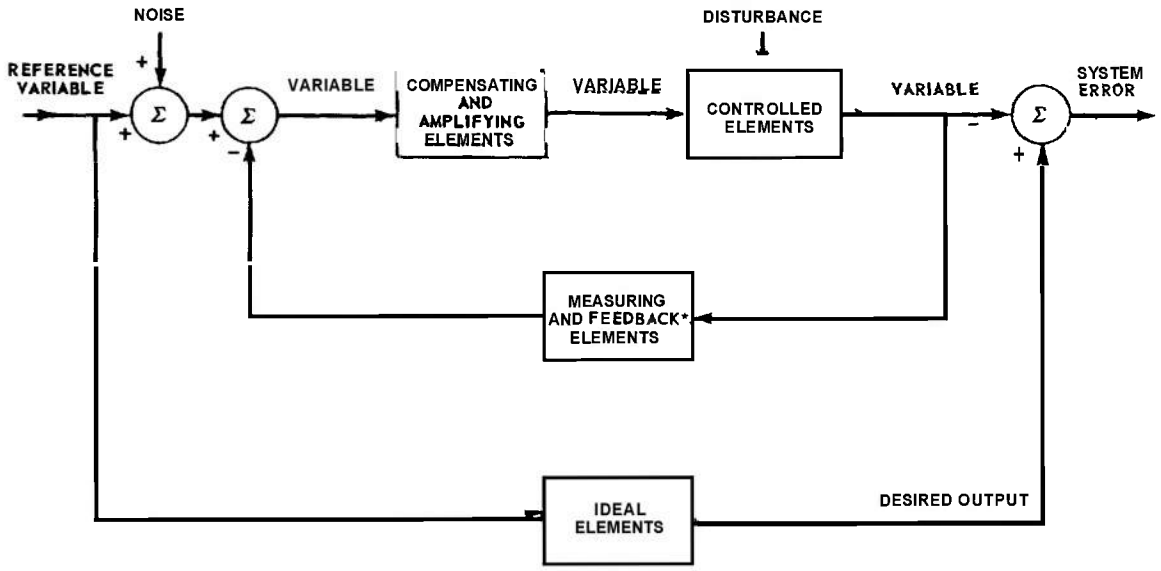
r = reference variable or input	u = disturbance or upset
n = noise	c = controlled **variable** or output
e = actuating variable	i = desired or ideal output
m = manipulated variable	y_e = system error

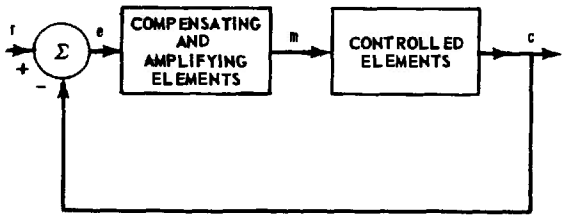
In many cases, the representation of Fig. 1-3 can be simplified. If the measuring and feedback elements are ideal and have no dynamic lag, it is possible to redraw the figure (see Ch. 3) so as to have no elements in the feedback path of the system.

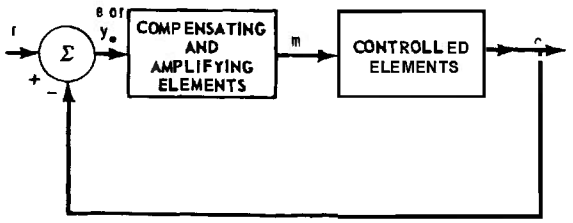
A system in which the unmodified controlled variable is fed back directly for comparison with the input is called a unity-feedback system. The main loop of a unity-feedback system is shown in Fig. 1-4.

If the ideal output of a system is the reference variable, the ideal elements are perfect. That is, the desired output is exactly equal to the reference input at all times. Many designs require that the output equal the input at all times although, strictly speaking, this is impossible with real components.

If a unity-feedback system is to have its desired output equal to its reference input, in the absence of noise, the system error must equal the actuating variable. A unity-feedback system of this type is often used initially in the process of design because of its simplicity. Such a system is shown in Fig. 1-5.







CHAPTER 2

DYNAMIC RESPONSE*

2-1 INTRODUCTION

The ***dynamic response*** of a component or system is the output response to an input that is a varying function of time. The ***steady-state response*** of a component or system is the output response to an input that is constant with time.

Paragraph 1-3 indicates that dynamic response is a basic determinant of system stability as well as an important element of system performance. All design theory for a feedback control system is centered on the study, analysis, and manipulation of the dynamic response characteristics of the system and of the components that are part of the system. Because of its fundamental importance, the dynamic response of any physical device or system is classified according to the nature of the input time variation that occurs. In some cases, the input time variation may be entirely artificial since it may not ordinarily occur in practice (for example, a sinusoidal signal). In other cases, the input variation may be one that is known to occur in practice (for example, a step change). In the former case, the artificial input time function is used primarily to facilitate analysis, design, and testing. In the latter case, the actual response of the system to the known input function is an important measure of performance which both the designer and user need to know in order to verify that the system meets the performance specifications. In either case, a clear understanding of the nature of the input and of the methods for finding the response to it are necessary for successful design.

2-2 LINEARIZATION

The basic tool used to describe the dynamic performance of a device is the set of differential equations that serve as a mathematical model for the actual physical device. Since quantitative techniques are imperative for analysis and design, a mathematical description is necessary. However, when going from the physical device to the differential equation model, one must resort to approximations if usable results are to be expected from a reasonable expenditure of time and effort. If the physical situation is such that it is possible to describe the device with a set of constant-coefficient linear differential equations to a high degree of accuracy, a wide variety of powerful tools are available to aid analysis. Even when the expected range of variation of the variables is such that the accuracy of approximation is partially lost when constant-coefficient linear differential equations are used, such a representation still serves a useful purpose. Although the representation above is inaccurate, it does provide a qualitative estimate of behavior which is still good enough to be of value to the designer for guiding testing procedures. Furthermore, if the designer artificially restricts the range

By L. A. Gould

of variation of the variables, he can obtain accurate results which apply to some, though not all, of the expected variations. From such a restriction, there results a partially accurate description that can at least be used to verify whether or not the device meets some of the performance specifications.

Although descriptions utilizing constant-coefficient linear differential equations predominate in feedback control system design, two classes of systems exist that do not lend themselves to such a description. Sampled- or pulsed-data systems are best described by variable-coefficient linear differential equations and are discussed in Ch. **9.** Contactor or relay servomechanisms cannot be described by linear equations at all, and one must resort to the nonlinear equations that describe these systems (see Ch. **10).** In addition, although the nonlinear properties of linear systems are ordinarily treated as secondary effects in the usual design procedure, they can, under certain circumstances, seriously affect performance. Such circumstances occur when the range of variation of the variables is wide or when the nonlinearity cannot be justifiably ignored. Secondary or incidental nonlinearities such as saturation and backlash are treated in Ch. 10.

Since linearization methods are the basis for most design work, it is important to understand the techniques that are used to establish a linear differential equation description of a device. *As a first step in a linearization procedure, appropriate assumptions are usually formulated based on a knowledge of the phhysics involved.* For example, in describing d-c machinery operating in an unsaturated region, the effect of hysteresis is often ignored and the normal magnetization curve of the steel is used in the analysis. The justification for such an approximation is based on the fact that, in many cases, the width of the hysteresis loop is small compared with the range of variation **of** magnetization encountered when using the device. In another situation, one may ignore the effect of backlash in a gear train for reasons analogous to those above.

When reasonable assumptions have eliminated many of the incidental nonlinearities of a device, one is often left with a performance description that is still nonlinear because of the curvature of the steady-state response (steady output as a function of a constant input) curves of the device. When this occurs, use is made of incremental techniques to produce the desired linear description.

The incremental linearization of a nonlinear characteristic (approximate representation of a nonlinear function by a linear function for small changes of the independent variable) is based on the Taylor series expansion of the function around a desired operating point. The deviation of the function from the operating point obtained from the approximate expansion of a function around a steady-state operating point is given by

$$\Delta f(x_1, x_2, \ldots, x_n) = C_{x_1}\Delta x_1 + C_{x2}\Delta x_2 + \ldots + C_{x_n}\Delta x_n \qquad \textbf{(2-1)}$$

$$\Delta f(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n) - f(x_{10}, x_{20}, \ldots, x_{n0}) \qquad \textbf{(2-2)}$$

where

$f(x_1, x_2, \ldots, x,)$ = function to be approximated

$(x_{10}, x_{20}, \ldots, x_{n0}$ = steady-state operating point

$$\Delta x_k = x_k - x_{k0} \qquad (k = 1, 2, \ldots, n)$$

$$C_{x_k} = \frac{\partial f}{\partial x_k}\Bigg|_{x_{10}, x_{20}, \ldots, x_{n0}} \qquad (k = 1, 2, \ldots, n)$$

In the approximation [Eq. **(2-1)],** the deviation or increment of the function from the operating point has been expressed as a linear function of the deviations (from the operating point) of its independent variables. The constant coefficients C_{x_k} are called the partial coefficients of f with respect to $x_1, x_2, \ldots, x_n$ at the operating point $(x_{10}, x_{20}, \ldots, x_{n0})$.

Example. A shunt d-c motor is speed-controlled by the vacuum-tube circuit shown in Fig. **2-1.** Assuming that hysteresis is negligible, the basic equations of the system are as follows:

$$E_p = E_p\,(E_g, I_f) \qquad \textbf{(2-3)}$$

$$If = \frac{E_s - E_p - N_f \frac{dF}{dt}}{R_f} \quad (2\text{-}4)$$

$$F = F(I_f) \quad (2\text{-}5)$$

$$L_a \frac{dI_a}{dt} + R_a I_a = E_s - E_b \quad (2\text{-}6)$$

$$E_b = k_e F N \quad (2\text{-}7)$$

$$M = k_m F I_a \quad (2\text{-}8)$$

$$M = J \frac{dN}{dt} + M_L \quad (2\text{-}9)$$

where

E_p = plate voltage
E_g = grid voltage
I_f = field current
E_s = supply voltage
E_b = motor back emf
L_a = armature inductance
R_a = armature resistance
N = motor shaft speed
k_e = motor back emf constant
N_f = number of field turns
F = field flux
R_f = field resistance
I_a = armature current
M = motor torque
k_m = motor torque constant
J = total moment of inertia
M_L = load torque

It is assumed that mechanical friction is negligible.

To linearize Eq. (2-3), let†

$$E_p \triangleq E_{po} + e_p = E_{po} + \Delta E_p \quad (2\text{-}10)$$

$$E_g \triangleq E_{go} + e_g = E_{go} + \Delta E_g \quad (2\text{-}11)$$

$$I_f \triangleq I_{fo} + i_f = I_{fo} + \Delta I_f \quad (2\text{-}12)$$

where E_{po}, E_{go}, and I_{fo} represent values at the steady-state operating point and e_p, e_g, and i_f represent the deviations of the values of E_p, E_g, and I_f from their corresponding steady-state operating-point values.

†Symbol $\triangleq$ is defined as "equals by definition".

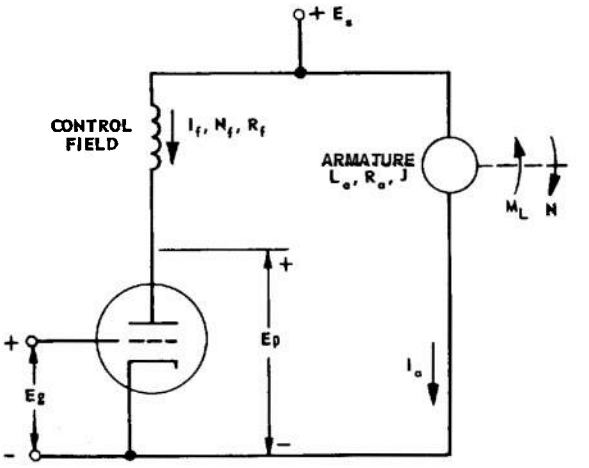


Fig. 2-1 Speed control of shunt d-c motor.

$$\text{Since } \Delta E_p = \left(\frac{\partial E_p}{\partial E_g}\right) \Delta E_g + \left(\frac{\partial E_p}{\partial I_f}\right) \Delta I_f$$

$$= \left(\frac{\partial E_p}{\partial E_g}\right) e_g + \left(\frac{\partial E_p}{\partial I_f}\right) i_f,$$

Eq. (2-10) becomes

$$E_p = E_{po} + (-\mu)\, e_g + (R_p)\, i_f \quad (2\text{-}13)$$

where

$$-\mu \triangleq \left.\frac{\partial E_p}{\partial E_g}\right|_{E_{go},\, I_{fo}} \quad \text{and} \quad R_p \triangleq \left.\frac{\partial E_p}{\partial I_f}\right|_{E_{go},\, I_{fo}}$$

As a result of the definitions above, the incremental linear approximation to Eq. (2-3) becomes

$$e_p = -\mu e_g + R_p i_f \quad (2\text{-}14)$$

where e_p, e_g, and i_f represent incremental quantities.

To linearize Eq. (2-5), let

$$F \triangleq F_o + f = F_o + \Delta F \quad (2\text{-}15)$$

Then, Eq. (2-15) can be written as

$$F = F_o + C_f i_f \quad (2\text{-}16)$$

where

$$C_f \triangleq \left.\frac{\partial F}{\partial I_f}\right|_{If^*}$$

Consequently, the incremental linear approximation to Eq. (2-5) becomes

$$f = C_f i_f \tag{2-17}$$

Substituting Eqs. (2-1o), (2-11), (2-12), (2-15), and (2-17) into Eq. (2-4) results in

$$I_{fo} + i_f = \frac{E_{so} + e_s - (E_{po} + e_p) - N_f \frac{d}{dt}(F_o + C_f i_f)}{R_f} \tag{2-18}$$

or

$$I_{,,} + i_f = \left(\frac{E_{so}}{R_f} = \frac{E_{po}}{R_f}\right) \left(\frac{e_s - e_p - N_f C_f \frac{di_f}{dt}}{R_f}\right) \tag{2-19}$$

where

$E_s \triangleq E_{so} + e_s$

E,, = steady-state value of E,

e, = increment in E,

It can be seen that the equation for the operating point of the field circuit is

$$I_{fo} = \frac{E_{so} - E_{po}}{R_f} \tag{2-20}$$

and the incremental equation for the field circuit is

$$i_f = \frac{e_s - e_p - N_f C_f \frac{di_f}{dt}}{R_f} \tag{2-21}$$

To linearize Eq. (2-7), let

$$E_b = E_{bo} + k_e N_o f + k_e F_o n \tag{2-22}$$

where

$$E_b \triangleq E_{bo} + e_b \tag{2-23}$$

$$N \triangleq N_o + n$$

Then, the incremental linear approximation for Eq. (2-7) becomes

$$e_b = k_e (N_o f + F_o n) \tag{2-24}$$

Substituting Eqs. (2-7), (2-17), (2-21), (2-23), and (2-24) into Eq. (2-6), and using a procedure analogous to that above, the equation for the operating point of the armature circuit becomes

$$I_{ao} = \frac{E_{so} - k_e F_o N_o}{R_a} \tag{2-25}$$

and the incremental equation for the armature circuit is

$$L_a \frac{di_a}{dt} + R_a i_a = e, - k_e N_o C_f i_f - k_e F_o n \tag{2-26}$$

where

$I_a \triangleq I_{ao} + i_a$

To linearize Eq. (2-8), use the incremental linear approximation

$$m = k_m F_o i_a + k_m I_{ao} f \tag{2-27}$$

where

$$M \triangleq M_o + m \tag{2-28}$$

By substituting Eqs. (2-23) and (2-28) into Eq. (2-9), the equation for the operating point of the mechanical circuit is

$$M_o = M_{Lo} \tag{2-29}$$

and the incremental equation of the mechanical circuit is

$$m = J \frac{dn}{dt} + m_L \tag{2-30}$$

where

$M_L \triangleq M_{Lo} + m_L$

From an examination of the analytical work above, it can be seen that the application of the linearizing technique produces a set of incremental equations that describe the behavoir of the system for deviations of the variables from the operating point of the system. In addition, the operating point is also defined by a set of algebraic equations.

Summarizing the operating-point equations:

$$I_{fo} = \frac{E_{so} - E_{po}}{R_f} \tag{2-20}$$

$$E_{po} \triangleq E_p\,(E_{go}, I_{fo}) \tag{2-31}$$

$$I_{ao} = \frac{E_{so} - k_e F_o N_o}{R_a} \tag{2-25}$$

$$F_o \triangleq F\,(I_{fo}) \tag{2-32}$$

$$M_o = M_{Lo} \tag{2-29}$$

$$M_o \triangleq k_m F_o I_{ao} \tag{2-33}$$

The operating-point equations can be solved for the unknowns I_{fo}, E_{po}, $I_{a,,}$ F_o, N_o, and M_o if the quantities $E_{s,,}$ $E_{g,,}$ and M_{Lo} are specified. This solution is usually done graphically because the steady-state characteristics of the tube [Eq. (2-31)] and the field structure [Eq. (2-32)] are presented as experimental curves.

Summarizing the incremental equations:

$$e_p = -\mu e_g + R_p i_f \tag{2-14}$$

$$i_f = \frac{e, - e_p - N_f C_f \frac{di_f}{dt}}{R_f} \tag{2-21}$$

$$L_a \frac{di_a}{dt} + R_a i_a = e, - k_c N_o C_f i_f - k_c F_o n \tag{2-26}$$

$$J \frac{dn}{dt} + m_L = k_m F_o i_a + k_m I_{ao} C_f i_f \tag{2-34}$$

The time-varying inputs to the system are the incremental quantities $e_{s,}$ $e_{g,}$ and m_L. If these quantities are known, the incremental equations can be solved for $e_{s,}$ i_f, $i_{a,}$ and n as functions of time.

2-3 TRANSIENT RESPONSE

The *transient response* of a system is the time variation of one or more of the system outputs following a sudden change in one or more of the system inputs or the derivatives or integrals of the system inputs. Often a transient input variation does not correspond to the actual input variations that a system might experience in practice. However, transient specifications of system performance are very commonly used and, as a result, the designer must know how to describe system behavior in terms of transient response. It can be shown (see Ch. 3) that the transient response of a *linear* system completely specifies the differential equations of the system and, therefore, can be used indirectly to find the response of the system to any type of input.

A given transient response must be referred to the type of input that caused it. Three commonly used transient test inputs are the *impulse,* the *step,* and the *ramp.*

A unit impulse can be conceived of as a time function that is infinite at $t = a$ and zero everywhere else. A unit impulse is defined by Eqs. (2-35) through (2-37), where $\delta_0(t-a)$ is a unit impulse function occurring at $t = a$.

$$\int_{-\infty}^{+\infty} \delta_0(t-a)\,dt = 1 \tag{2-35}$$

$$\int_{-m}^{+\infty} \delta_0(t-a) f(t)\,dt = f(a) \tag{2-36}$$

$$\delta_0(t-a) = 0, t > a \text{ and } t < a \tag{2-37}$$

The unit step function $\delta_{-1}(t-a)$ is merely the integral of the unit impulse $\delta_0(t-a)$. The unit step is defined by Eqs. (2-38) and (2-39), where $\delta_{-1}(t-a)$ is a unit step occurring at $t = a$.

$$\delta_{-1}(t-a) = \int_{-\infty}^{t} \delta_0(x-a)\,dx \tag{2-38}$$

$$\delta_{-1}(t-a) = \begin{cases} 0, t < a \\ 1, t > a \end{cases} \tag{2-39}$$

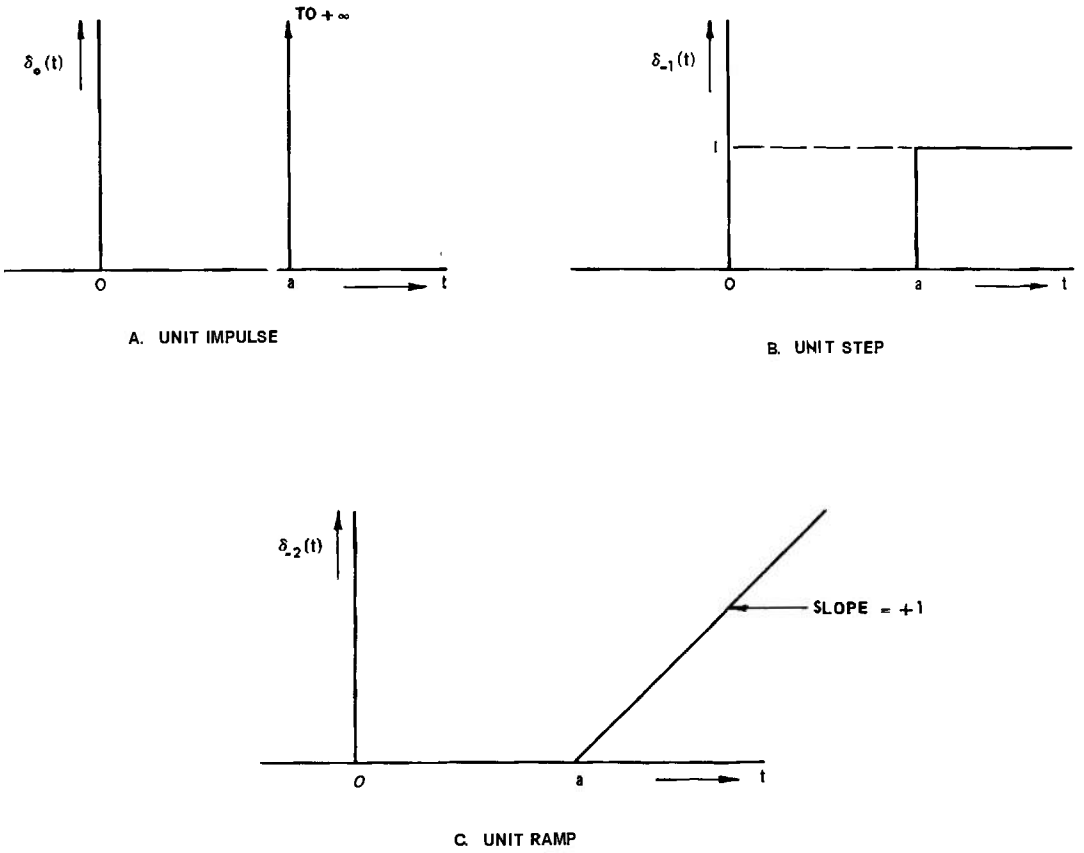


Fig. 2-2 Transient input functions.

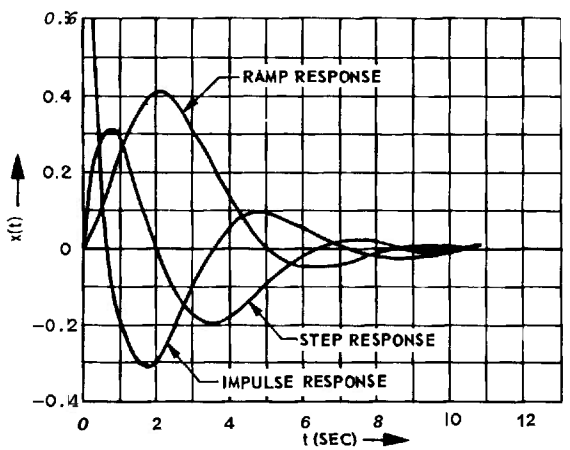


Fig. 2-3 Transient responses for system $\frac{d^3x}{dt^3} + 2\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + x = \frac{d^2y}{dt^2}$.

The unit ramp function $\delta_{-2}(t-a)$ is the integral of the unit step $\delta_{-1}(t-a)$. The unit ramp is defined by Eqs. (2-40) and (2-41), where $\delta_{-2}(t-a)$ is a unit ramp occurring at $t=a$.

$$\delta_{-2}(t-a) = \int_{-\infty}^{t} \delta_{-1}(x-a)\,dx \qquad (2\text{-}40)$$

$$\delta_{-2}(t-a) = \begin{cases} 0, & t < a \\ t, & t > a \end{cases} \qquad (2\text{-}41)$$

The impulse, step, and ramp functions are shown in Fig. 2-2. It should be noted that these functions are equal to zero for all $t < a$ and that they are discontinuous or one or more of their derivatives are discontinuous at the instant of occurrence. Clearly, the response of a physical system to any one of these inputs will be zero before the input occurs if the system is assumed to be initially at rest.

Example. A system is described by the equation

$$\frac{d^3x}{dt^3} + 2\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + x = \frac{d^2y}{dt^2} \qquad (2\text{-}42)$$

where y is the input and x is the output. The output transient responses as functions of time are shown in Fig. 2-3 when the input is a unit impulse, a unit step, and a unit ramp, each occurring at $t = 0$. The specified initial conditions for $t \leq 0$ are $x = 0$, $dx/dt = 0$, and $d^2x/dt^2 = 0$.

It should be noted that, although the curves in Fig. 2-3 are different, they represent the same information about system behavior provided that the input associated with each curve is known.

2-4 FREQUENCY RESPONSE

The *frequency response* of a system is the variation of the output to an input which is a constant-amplitude variable-frequency sinusoid. Frequency response is usually of interest in the linear case but does have application in the nonlinear case (see Ch. 10).

In the case of a linear system, a sinusoidal input produces a sinusoidal output of the same frequency as the input. The frequency response of a linear system is therefore completely described by the ratio of the output amplitude to the input amplitude and by the phase angle of the output relative to the phase angle of the input, both expressed as functions of frequency.

The frequency response of a system is usually presented in three ways: by a polar plot of the tip of the vector $A(\omega)\, e^{j\phi(\omega)}$ with frequency ω as a parameter ($j = \sqrt{-1}$); by separate plots of $10 \log_{10} A(\omega)$ and $\phi(\omega)$ versus frequency ω; and by the gain-phase plot of $10 \log_{10} A(\omega)$ versus $\phi(\omega)$ with frequency ω as a parameter. $A(\omega)$ is the amplitude ratio of output to input and $\phi(\omega)$ is the output phase angle minus the input phase angle. One could also plot $20 \log_{10} A(\omega)$ as is done in the literature in many places, but there is little to be gained by using the factor of two.

The frequency response of a system is useful primarily because of the many theoretical simplifications that are possible when it is used as an analytical and design tool. Just as transient inputs rarely occur in practice, so do sinusoidal inputs almost never occur in practice. Nevertheless, both methods of describing dynamic response are useful in analysis and design.

Since frequency response and transient response are directly related to the differential equation of a system, they contain the same information about system behavior. These two methods of describing dynamic response are merely different approaches to the same end. Both have a useful function to perform in designing control systems. Techniques for correlating the frequency response and transient response of a system are presented in Chs. 3 and 7.

2-5 FORCED RESPONSE

The *forced response* of a system is the time response of an output of the system to an arbitrary, but completely defined, variation of one of the system inputs. Forced response is distinguished from transient response in that the input variation associated with the forced response of a system is considered as a continuous time function with no discontinuities in any of its derivatives. A sinusoidal input is a special case of a forcing input which is isolated for special attention because of its theoretical importance.

A typical example of an arbitrary forcing input is the angle of the line-of-sight from a radar antenna to a target that is moving at constant velocity in a straight line (see Fig. 2-4). Such a course is known as a ***straight-line crossing*** course. The angle of the line-of-sight θ in this case is given by

$$\theta(t) = \tan^{-1} \frac{V}{R} t \qquad (2\text{-}43)$$

where

V = target velocity

R = minimum target range

The inverse tangent function in Eq. (2-43) and all its derivatives are continuous for all t.

Many design problems have input specifications involving arbitrary forcing functions that cannot be adequately described by discontinuous functions. The techniques for determining the response of a system to these functions are discussed in Pars. 3-3, 3-7, and 7-2.

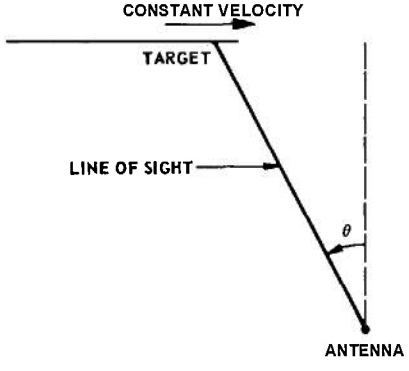


Fig. 2-4 Straight-line crossing course.

2-6 STOCHASTIC INPUTS

A *stochastic process* is one in which there is an element of chance. In many situations, the input to a system is not completely predictable and cannot be described by a mathematical function, either analytically or graphically. The term "random process" is often used to describe such a situation, but it is not an accurate term since a process can often consist of a combination of a completely predictable portion together with a purely random portion. It is evident that the signals in a feedback control system are more often stochastic than predictable in nature, particularly when the effect of the ever-present noise is considered. A typical example of a stochastic process is a radar signal corrupted by noise.

Since a degree of uncertainty exists if one attempts to determine the value of a stochastic signal at a given instant of time, probability density functions and other statistical characterizations such as the average value, the root-mean-square (rms) value, and the correlation function are used to describe the signal. It is useful to think of a stochastic signal as a member of a family of signals, each generated by an identical process. Such a family of signals is called an *ensemble,* and the statistical characterizations of the stochastic process are related to the ensemble of signals rather than to a particular member of the ensemble. Determination of the response of a system to a stochastic input does not yield a function of time, but rather a

statistical characterization of the output signal ensemble.

Stochastic signals are separated into two classes. If the statistical behavior of the process that generates the ensemble is independent of time, the process is ***stationary.*** A ***nonstationary*** process is one whose statistics vary with time. In most situations involving stochastic processes, the signals generated are non-stationary. It is useful, however, to treat practical processes as stationary if the variation of the statistics with time is small over the useful life of the system. A typical example is the noise generated in a vacuum tube. As the tube ages, the statistical character of the noise changes. If the period of use of the tube is short compared with its expected life, then the noise generated by the tube can be considered as a stationary process.

If a stochastic process is stationary, it is possible to use a single member of the ensemble of the process to determine the statistics of the process. For example, if the average value of a signal is sought and the process generating the signal is known to be stationary, the average value can be found in two ways. In the first way, the average value is computed by taking the time average for a single member of the ensemble. In the second way, the average value is computed by taking the values of all the members of the ensemble at a particular instant of time and averaging these values. The latter average is called the ***ensemble average.*** Since the process is known to be stationary, both averages are independent of time. That both averages are identical has not been proven as yet, but their identity seems plausible if one accepts the assumption, the so-called ***ergodic hypothesis,*** that ensemble averages and time averages are identical for a stationary process. The various ways to characterize stochastic signals and the response of a system to a stochastic signal are discussed in Par. 3-8.

CHAPTER 3

METHODS OF DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS*

3-1 THE DIFFERENTIAL EQUATIONS

As discussed in Ch. 2, any design procedure is based on the differential equations that serve as the mathematical model for the physical system. This chapter deals with methods of determining the dynamic response of physical systems from the differential equations that describe them. The type of response sought depends upon several factors: the specifications of the system; the design procedure adopted; and the limitations imposed by test conditions encountered when seeking experimental verification of the design performance.

Differential equations may be classified as follows:

(a) Linear differential equations with constant coefficients

(b) Linear differential equations with time-varying coefficients

(c) Nonlinear differential equations

Of the three classes, constant-coefficient linear differential equations are, by far, the most widely used and the best understood. The subject matter of this chapter is focused exclusively on methods of solving equations in this class. Chapter 9 deals with time-variable linear differential equations of a specific type that have a wide application. Chapter 10 considers nonlinear equations and some of the techniques for treating them.

The general form of a linear differential equation with constant coefficients is

$$\sum_{i=0}^{n} a_i \frac{d^i x}{dt^i} = \sum_{j=0}^{m} b_j \frac{d^j y}{dt^j} \qquad (3\text{-}1)$$

where the a's and b's are the constant coefficients, $x(t)$ is the response function, and $y(t)$ is the input function. The equation is linear because the response to a sum of component input functions equals the sum of the responses to each of the component input functions. The highest-order derivative of the response x present in the equation is called the *order* of the equation. Thus, Eq. **(3-1)** is an equation of the n^{th} order. The information necessary for a solution of the equation is a statement of the initial value of the response and the initial values of its first n-1 derivatives, as well as the value of the input $y(t)$. The response can be separated into two parts —a general or homogeneous solution, and a particular solution. The complete solution of the differential equation is the sum of the general solution and the particular solution. The general solution always has the form of a sum of exponentials with real and complex arguments; the particular solution has the same form as the input or a sum of the input and its derivatives. The general solution is often called the *force-free* or *transient* solution; the particular solution is called the *forced* or *steady* solution. Each term in the transient solution is called a *normal response mode* or *characteristic* of the equation.

The complete solution of a linear differential equation is given by

$$x(t) = x_p(t) + \sum_{k=1}^{n} A_k e^{p_k t} \qquad (3\text{-}2)$$

where $x_p(t)$ is the particular solution, p_k is a root of the equation, and A_k is a constant-amplitude coefficientof a response mode. The

*By L. A. Gould

root p_k is a function only of the coefficients a_i whereas A, is a function of the a's, b's, and $y(t)$ [Eq. (3-1)]. The quantities A, and p_k are generally complex numbers that occur in conjugate pairs if the coefficients a_i and b_j and the input function $y(t)$ are real.

The term *root* is applied to each of the p_k's because these numbers can be found from the differential equation by treating the differentiating operator d/dt as a real variable, replacing it by p for convenience, and setting $y(t)$ equal to zero. The algebraic equation resulting from such substitutions in Eq. (3-1) is

$$\sum_{i=0}^{n} a_i p^i = 0 \tag{3-3}$$

This equation is known as the *characteristic* equation. The roots of Eq. (3-3), when determined, give the p_k's of the normal response modes of Eq. (3-2).

The classical procedure for solving constant-coefficient linear differential equations is covered in many textbooks.[1,2,3,4] More powerful tools for treating differential equations, such as Laplace and Fourier transforms, are presented in Par. 3-4. For situations where the input is sinusoidal or stochastic, additional special techniques are used. These are discussed later in the text.

3-2 FACTORING AND CHARACTERISTIC PARAMETERS OF RESPONSE MODES

3-2.1 FACTORING

In most cases, the solution of a linear differential equation requires the determination of the roots of the characteristic equation [Eq. (3-3)]. Unfortunately, if the order of the equation is high, the process of factoring the equation to find the roots becomes extremely tedious. For such cases, special techniques have been developed (see Pars. 4-4, 5-7, and 7-1). This section covers some general factoring procedures applicable to any algebraic equation. In addition, the characteristics of first- and second-order equations are discussed and graphical methods for determining the roots of third- and fourth-order equations are presented.

The factoring of rational polynomials is covered by many authors.[5,6,7,8,9,10,33] The method presented here is one that is very convenient.

The general algebraic equation can be written as

$$f(p) = p^n + c_{n-1}p^{n-1} + \ldots\ldots + c_1 p + c_0 = 0 \tag{3-4}$$

If the order n of the equation is odd, one or more real roots must exist. The real root (or roots) can be determined graphically by plotting $f(p)$ versus p and noting the zero-crossing(s) of $f(p)$, or analytically by using Horner's method of synthetic division, with the first trial divisor being

$$(p + p_1) = p + \frac{c_0}{c_1} \tag{3-5}$$

If the equation is reduced to one of even order, Lin's method[9] can be used. This involves choosing the tria! divisor

$$g_1(p) = p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2} \tag{3-6}$$

Next, $f(p)$ is divided by $g_1(p)$ as follows:

$$\begin{array}{r} p^{n-2} + \ldots \\ p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2} \;\overline{)\; p^n + c_{n-1}p^{n-1} + \ldots + c_1 p + c_0} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ c'_2 p^2 + c'_1 p + c_0 \\ \underline{c'_2 p^2 + c'_1 p + c'_0} \\ \text{remainder} \end{array} \tag{3-7}$$

If the remainder of Eq. (3-7) is not negligible, a new trial divisor is chosen such that

$$g_2(p) = p^2 + \frac{c'_1}{c'_2}p + \frac{c_0}{c'_2} \qquad (3\text{-}8)$$

Next, $f(p)$ is divided by the new divisor $g_2(p)$, etc., and the process is continued until the remainder is negligible. The last divisor is then a factor of the original equation. Then, the quotient (of $f(p)$ and this last divisor] is treated in an identical manner, and the process is repeated until f(1) is factored into quadratic factors whose roots can be determined directly.

Example. Find the roots of the algebraic equation

$$f(p) = p^4 + 10.65p^3 + 89.0p^2 + 15.50p + 27.0 = 0 \qquad (3\text{-}9)$$

Solution. The first trial divisor is

$$g_1(p) = p^2 + \frac{15.50p}{89.0} + \frac{27.0}{89.0} = p^2 + 0.1742p + 0.303 \qquad (3\text{-}10)$$

Dividing $f(p)$ by $g_1(p)$ produces

$$\begin{array}{r|l}
 & p^2 + 10.48p + 86.9 \\
\hline
p^2 + 0.1742p + 0.303 & p^4 + 10.65p^3 + 89.0p^2 + 15.50p + 27.0 \\
 & \underline{p^4 + 0.17p^3 + 0.3p^2} \\
 & \quad 10.48p^3 + 88.7p^2 + 15.50p \\
 & \quad \underline{10.48p^3 + 1.8p^2 + 3.18p} \\
\text{second trial divisor} \longrightarrow & \quad 86.9p^2 + 12.32p + 27.0 \\
 & \quad \underline{86.9p^2 + 15.14p + 26.3} \\
\text{remainder} \longrightarrow & \quad -2.82p + 0.7
\end{array}$$

The second trial divisor is

$$g_2(p) = p^2 + \frac{12.32}{86.9}p + \frac{27.0}{86.9} = p^2 + 0.1418p + 0.311 \tag{3-11}$$

Dividing $f(p)$ by $g_2(p)$ yields

$$\begin{array}{r|l}
 & p^2 + 10.51p + 87.2 \\
\hline
p^2 + 0.1418p + 0.311 & p^4 + 10.65p^3 + 89.0p^2 + 15.50p + 27.0 \\
 & \underline{p^4 + 0.14p^3 + 0.3p^2} \\
 & 10.51p^3 + 88.7p^2 + 15.50p \\
 & \underline{10.51p^3 + 1.5p^2 + 3.27p} \\
\text{third trial divisor} \longrightarrow & 87.2p^2 + 12.23p + 27.0 \\
 & \underline{87.2p^2 + 12.36p + 27.1} \\
\text{remainder} \longrightarrow & -0.13p - 0.1
\end{array}$$

The third trial divisor is

$$g_3(p) = p^2 + \frac{12.23}{87.2}p + \frac{27.0}{87.2} = p^2 + 0.1403p + 0.310 \tag{3-12}$$

Since $g_3(p)$ leaves essentially no remainder, the resulting quadratic factors of $f(p)$ are

$$f(p) = (p^2 + 0.1403p + 0.310)\ (p^2 + 10.51p + 87.2) \tag{3-13}$$

Factoring the two quadratics in Eq. (3-13), the roots of Eq. (3-9) are found to be

$$p_1, p_2 = -0.0702 \pm j\,0.552 \text{ and } p_3, p_4 = -5.26 \pm j\,7.72 \tag{3-14}$$

3-2.2 CHARACTERISTIC PARAMETERS OF RESPONSE MODES

3-2.3 First Order: $(p + c_0) = 0$

The response mode corresponding to the first-order factor is of the form

$$x(t) = Ae^{-c_0 t} \tag{3-15}$$

The reciprocal of c_0 is called the ***time constant*** of the response mode.

Useful plots of exponential functions are presented in Fig. 3-1.

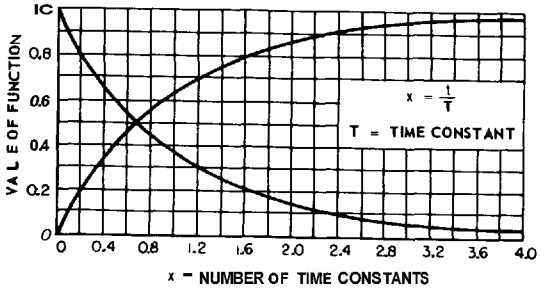


Fig. 3-7 Exponential functions e^{-x} *and* $1 - e^{-x}$.

3-2.4 Second Order: $p^2 + c_1 p + c_0 = 0$

The second-order equation can be rewritten in the form

$$p^2 + 2\zeta\omega_n p + \omega_n^2 = 0 \qquad (3\text{-}16)$$

where

$\zeta = \dfrac{c_1}{2\sqrt{c_0}}$ = damping ratio

$\omega_n = \sqrt{c_0}$ = undamped natural frequency

If $\zeta < 1$, the response mode corresponding to the second-order factor is of the form

$$x(t) = Ae^{-\zeta\omega_n t}\cos(\omega_d t + \phi) \qquad (3\text{-}17)$$

where

$\omega_d = \omega_n\sqrt{1-\zeta^2}$ = damped frequency of transient oscillation

If $\zeta \geqq 1$, the second-order factor can be factored into two first-order factors so that the response consists of two first-order modes.

3-2.5 Third Order[11]

$$p^3 + c_2 p^2 + c_1 p + c_0 = 0$$

By making the substitution

$$p = c_0^{1/3}\lambda$$

the third-order equation is reduced to

$$\lambda^3 + \alpha_2\lambda^2 + \alpha_1\lambda + 1 = 0 \qquad (3\text{-}18)$$

where

$$\alpha_2 = c_2/c_0^{1/3} \text{ and } \alpha_1 = c_1/c_0^{2/3}$$

The reduced equation can be factored as follows:

$$\lambda^3 + \alpha_2\lambda^2 + \alpha_1\lambda + 1 = (\lambda + 1/\omega_r^2)(\lambda^2 + 2\zeta\omega_r\lambda + \omega_r^2) \qquad (3\text{-}19)$$

where

$$\alpha_2 = 2\zeta\omega_r + \frac{1}{\omega_r^2}$$

$$\alpha_1 = \frac{2\zeta}{\omega_r} + \omega_r^2$$

ω_r = a reference frequency

Figure 3-2 shows plots of ω_r versus ζ for constant values of α_2 and α_1. Figure 3-3 shows plots of α_2 versus α_1 for constant values of ζ and ω_r. From these charts and the third-order equations, the roots of the cubic can be determined.

3-2.6 Fourth Order[11]:

$$p^4 + c_3 p^3 + c_2 p^2 + c_1 p + c_0 = 0$$

By making the substitution

$$p = c_0^{1/4}\lambda \qquad (3\text{-}20)$$

the fourth-order equation is reduced to

$$\lambda^4 + \alpha_3\lambda^3 + \alpha_2\lambda^2 + \alpha_1\lambda + 1 = 0 \qquad (3\text{-}21)$$

where

$$\alpha_3 = c_3/c_0^{1/4}$$

$$\alpha_2 = c_2/c_0^{1/2}$$

$$\alpha_1 = c_1/c_0^{3/4}$$

The reduced equation can be factored into the form

$$(\lambda^2 + 2\zeta_1\omega_{r_1}\lambda + \omega_{r_1}^2)(\lambda^2 + 2\zeta_2\omega_{r_2}\lambda + \omega_{r_2}^2) = 0 \qquad (3\text{-}22)$$

or, alternatively

$$(\lambda^2 + 2\zeta_r\omega_r\lambda + \omega_r^2)[\lambda^2 + 2(\zeta_r\rho_\zeta)(\omega_r\rho\omega)\lambda + (\omega_r\rho_\omega)^2] = 0 \qquad (3\text{-}23)$$

where the symbols are defined as follows:

$\omega_r \triangleq \omega_{r_1}$ = dimensionless natural frequency of reference component

$\zeta_r \triangleq \zeta_1$ = damping ratio of reference component

$\rho_\omega \triangleq \omega_{r_2}/\omega_{r_1}$ = ratio of undamped natural frequencies of components

$\rho_\zeta \triangleq \zeta_2/\zeta_1$ = ratio of damping ratios of components

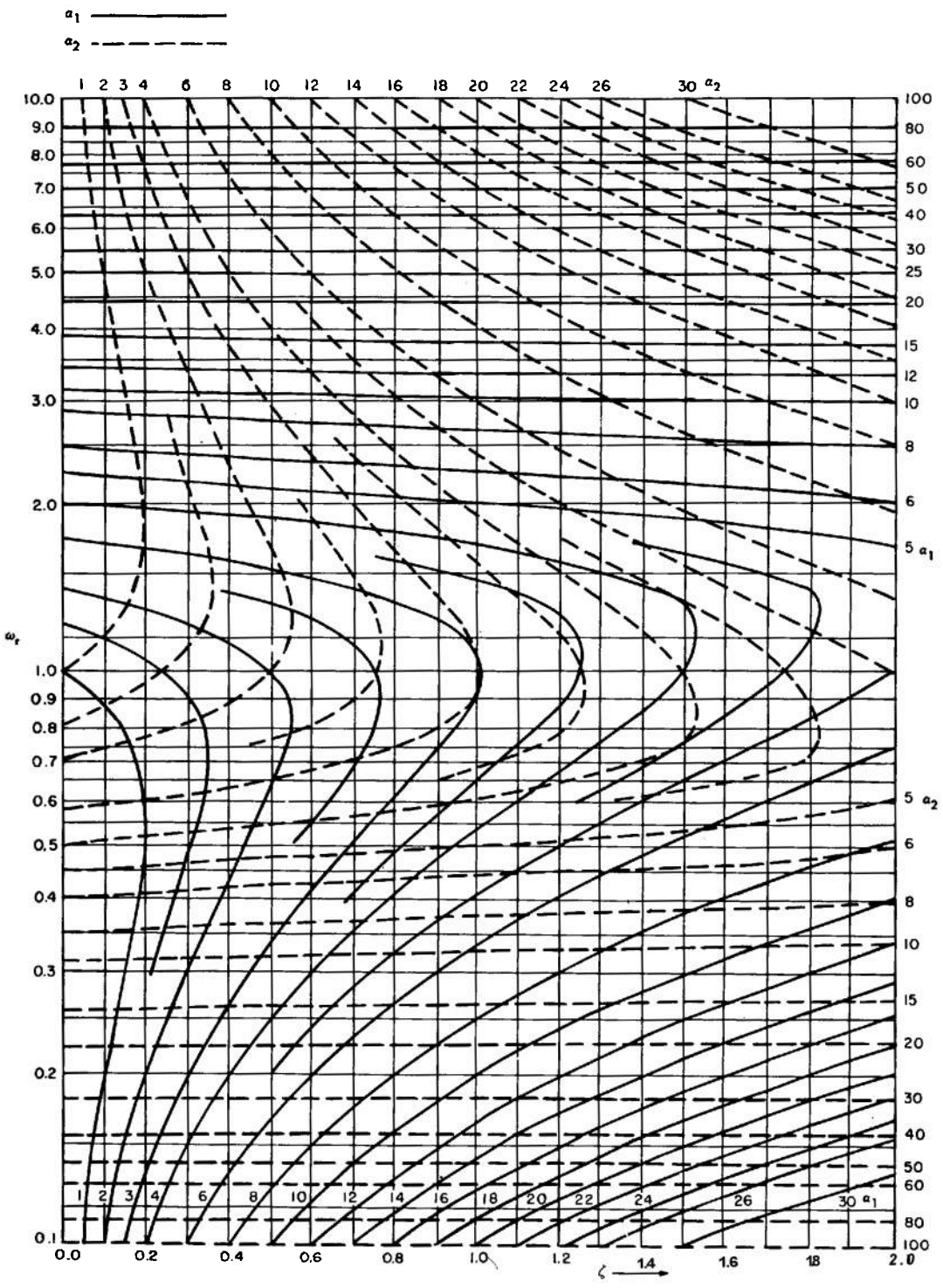


Fig. 3-2 Cubic chart.

By permission from "Solution of the Cubic Equations and the Cubic Charts", by L. W. Evans, bound with "Transient Behavior and Design of Servomechanisms". by G. S. Brown, 1943. Massachusetts Institute of Technology.

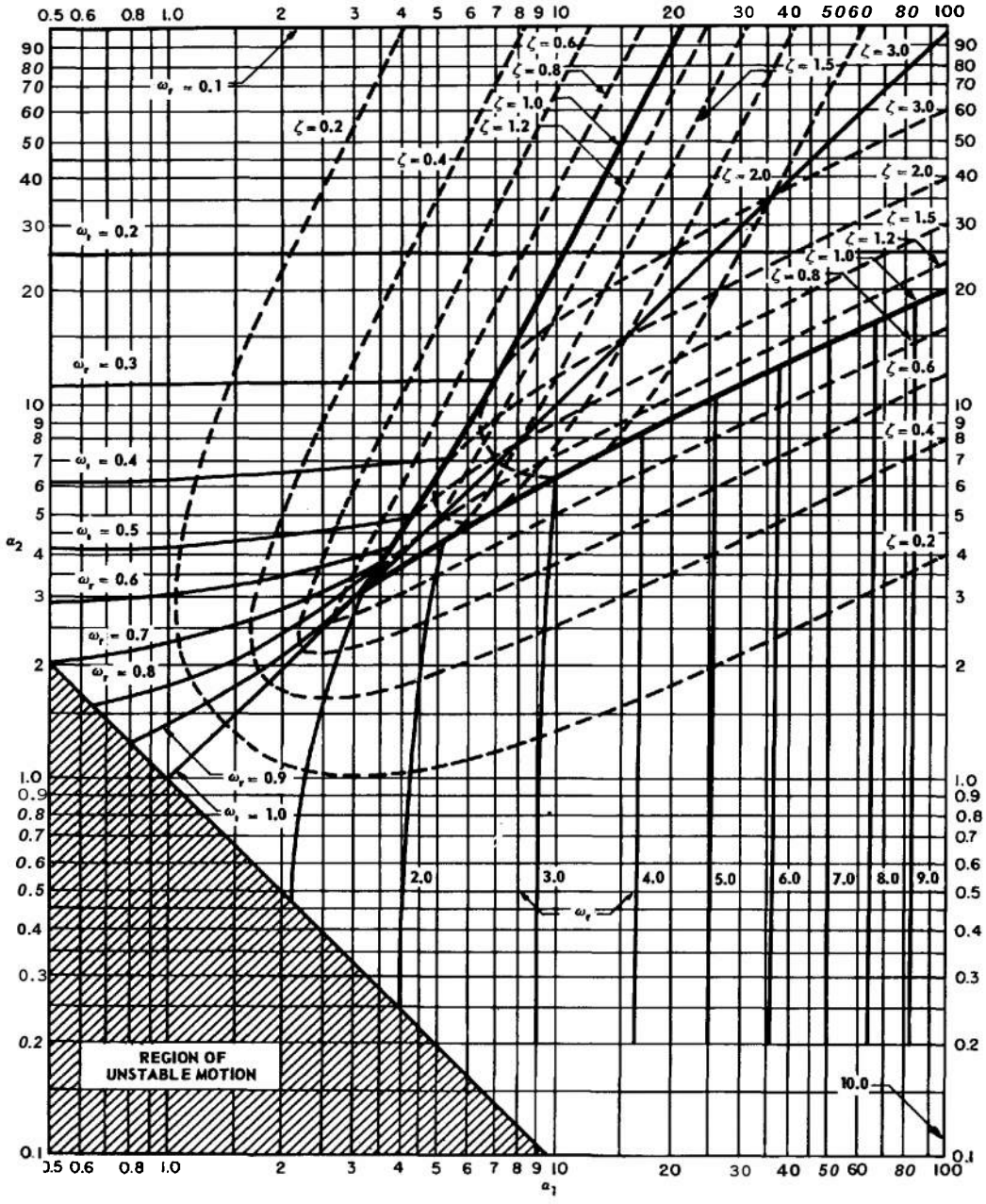


Fig. 3-3 *Cubic chart.*

By permission from "Solution of the Cubic Equations and the Cubic Charts", by L. W. Evans, bound with "Transient Behavior and Design of Servomechanisms", by G. S. Brown, 1943, Massachusetts Institute of Technology.

CHART I FOR EQUATION $\lambda^4 + a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + 1 = 0$

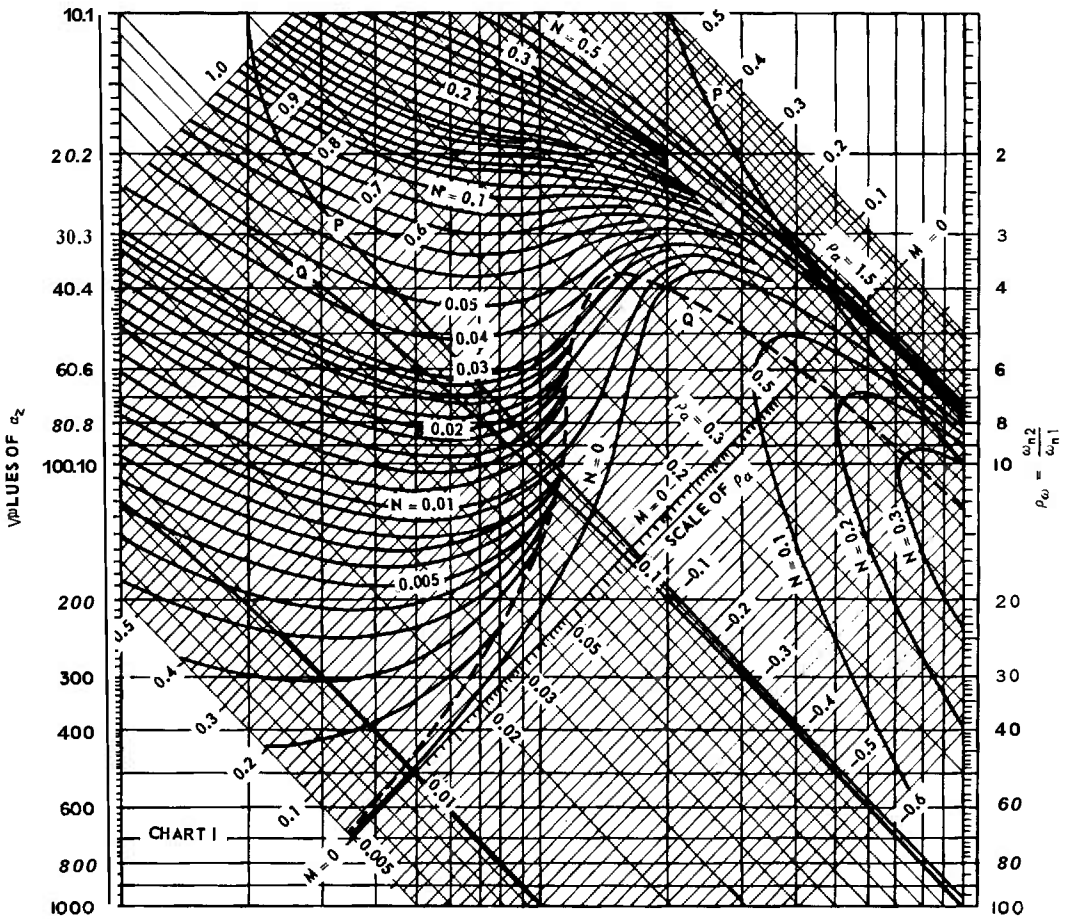


Fig. 3-4 Quartic chart. (Sheet I **of** *2)*

By permission from "Servomechanisms", by Y. J. Liu, bound with "Transient Behavior and Design of Servomechanisms". by G. S. Brown, 1943, Massachusetts Institute of Technology.

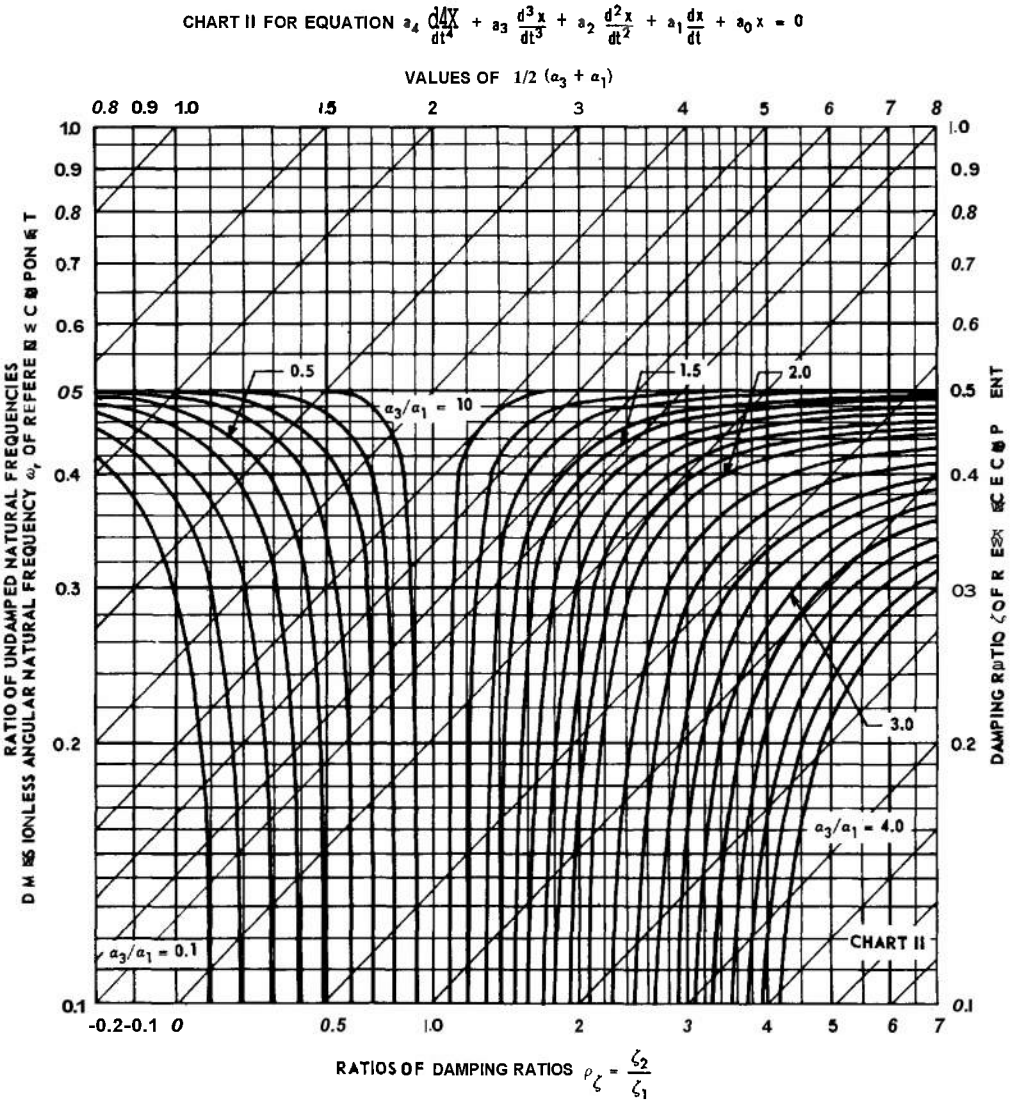


Fig. 3-4 Quartic Chart. (Sheet 2 of 2)

To find the parameters defined by Eqs. (3-22) and (3-23), the following procedure is used:

(a) Determine the quantities

$$M \triangleq \frac{\alpha_3\alpha_{1-4}}{\alpha_2^2}$$

$$N \triangleq \frac{\alpha_3^2 + \alpha_1^2 - 4\alpha_2}{\alpha_2^3}$$

$$\frac{\alpha_3}{\alpha_1} \text{ and } \frac{1}{2}(\alpha_3 + \alpha_1)$$

(b) Stability can be determined from Routh's criterion (Par. 4-2).

(c) The quartic chart is shown in Fig. 3-4. A sketch of the construction that is used to find ρ_ω, ω_r, ρ_ζ, and ζ_r is shown in Fig. **3-5.** Referring to these figures, the determination of the parameters of the factored quartic [Eq. (3-22)] is given by the procedure below.

(d) Locate intersection 3a of the particular pair of M and N values on Chart I. Draw a line through 3a, parallel to the 135°-inclined lines, until it intersects the 45"-inclined scale. The intersection on this scale gives ρ_α, where

$$\rho_\alpha \triangleq \frac{1}{\alpha_2}\left(\rho_\omega + \frac{1}{\rho_\omega}\right)$$

(e) From the particular α_2 value on the left-hand scale, draw a horizontal line until it meets the particular 135°-line found in step (d) at point 3b.

(f) From intersection 3b, draw a vertical line that intersects curve P at 3P and curve Q at 3Q.

(g) A horizontal line drawn through 3P intersects the immediate right scale of ordinates at 3d giving the value of ρ_ω, and the next right (left-hand scale of ordinates of Chart 11) at 3d' giving the value of ω_r.

(h) A horizontal line drawn through 3Q on Chart I intersects the particular curve of α_3/α_1 on Chart II at point 4a. The lower abscissa of 4a gives the value of ρ_ζ.

(i) Through 4a, draw a 45°-inclined line until it intersects a vertical line corresponding to the particular value of $(\alpha_3 + \alpha_1)/2$ at point 4b. A horizontal line drawn through 4b intersects the extreme right-hand scale of Chart II giving the value of ζ_r.

(j) When ρ_α is obtained in step (d), the following equations can be used as an alternate method of finding ρ_ω, ω_r, ρ_ζ, and ζ_r:

$$\rho_\omega = \frac{1}{2}\left[\alpha_2\rho_\alpha + \sqrt{(\alpha_2\rho_\alpha)^2 - 4}\right] \qquad (3\text{-}24)$$

$$\omega_r = \frac{1}{\sqrt{\rho_\omega}} \qquad (3\text{-}25)$$

$$\rho_\zeta = \frac{\rho_\omega\left(\frac{\alpha_3}{\alpha_1}\right) - 1}{\rho_\omega - \left(\frac{\alpha_3}{\alpha_1}\right)} \qquad (3\text{-}26)$$

$$\zeta_r = \frac{\frac{1}{2}(\alpha_3 + \alpha_1)}{(1 + \rho_\zeta)\left[\sqrt{\rho_\omega} + \frac{1}{\sqrt{\rho\omega}}\right]} \qquad (3\text{-}27)$$

$$\zeta_r = \frac{1}{2}\sqrt{\frac{\alpha_2(1 - \rho_\alpha)}{\rho_\zeta}} \qquad (3\text{-}28)$$

$$\zeta_r = \frac{\alpha_3}{2\left[\frac{1}{\sqrt{\rho_\omega}} + \rho_\zeta\sqrt{\rho_\omega}\right]} \qquad (3\text{-}29)$$

$$\zeta_r = \frac{\alpha_1}{2\left[\frac{\rho_\zeta}{\sqrt{\rho_\omega}} + \sqrt{\rho_\omega}\right]} \qquad (3\text{-}30)$$

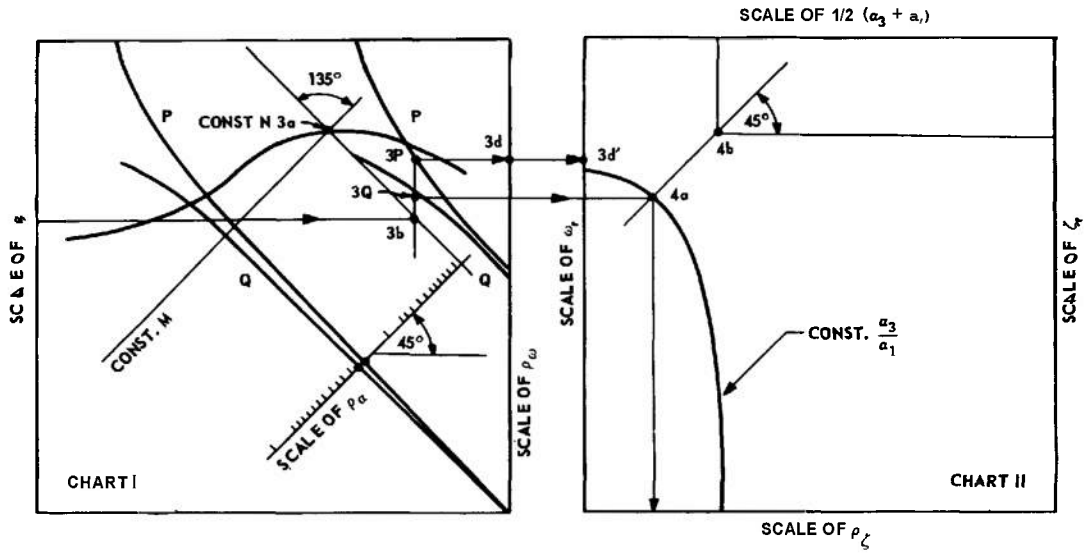


Fig. 3-5 Sketch of the quartic chart.

By permission from "Servomechanisms", by Y. J. Liu, bound with "Transient Behavior and Design of Servomechanisms", by G. S. Brown, 1943. Massachusetts Institute of Technology.

3-3 THE CONVOLUTION INTEGRAL[12]

The output time response of any linear system to an arbitrary input can be found by means of the *convolution* (superposition) integral. If $y(t)$ is the input, $x(t)$ the output, and w(t) the impulse response of the system, then the output x can be found by evaluating the convolution integral

$$x(t) = \int_{-\infty}^{+\infty} dt_1\, w(t_1)\, y(t - t_1) \qquad (3\text{-}31)$$

or

$$x(t) = \int_{-\infty}^{+\infty} dt_1\, w(t - t_1)\, y(t_1) \qquad (3\text{-}32)$$

This integral applies in every case and is useful for graphical time-domain studies of system performance. In many situations, however, evaluation of the convolution integral is tedious, so more refined procedures are used (see Pars. 3-4, 3-6, 3-7, 3-8, and 3-9).

If the system being studied is a physical system, then

$$w(t) = 0 \text{ for } t < 0 \qquad (3\text{-}33)$$

The convolution integral then reduces to

$$x(t) = \int_{0}^{+\infty} dt_1\, w(t_1)\, y(t - t_1) \qquad (3\text{-}34)$$

or

$$x(t) = \int_{-\infty}^{t} dt_1\, w(t - t_1)\, y(t_1) \qquad (3\text{-}35)$$

If, as often occurs, $y(t)$ and w(t) are both zero for $t < 0$, then the convolution integral reduces to

$$x(t) = \int_{0}^{t} dt_1\, w(t_1)\, y(t - t_1) \qquad (3\text{-}36)$$

or

$$x(t) = \int_{0}^{t} dt_1\, w(t - t_1)\, y(t_1) \qquad (3\text{-}37)$$

3-4 LAPLACE AND FOURIER TRANSFORMS[12,13,14,15,16,17,18,19]

3-4.1 GENERAL

Laplace and Fourier transforms are aids for solving differential equations and introduce properties of system performance that enhance the designer's understanding and simplify his task.

The *Fourier transform* of a function and its inverse are defined as follows:

$$\mathcal{F}[f(t)] \triangleq F(s) \triangleq \int_{-\infty}^{+\infty} dt\, e^{-st} f(t) \quad \text{[Direct]} \tag{3-38}$$

$$\mathcal{F}^{-1}[F(s)] \triangleq f(t) \triangleq \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} ds\, e^{st} F(s) \quad \text{[Inverse]} \tag{3-39}$$

where $s = \sigma + j\omega$

The *Laplace transform* of a function and its inverse are defined as follows:

$$\mathcal{L}[f(t)] \triangleq F(s) \triangleq \int_{0}^{\infty} dt\, e^{-st} f(t) \quad \text{[Direct]} \tag{3-40}$$

$$\mathcal{L}^{-1}[F(s)] \triangleq f(t) \triangleq \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} ds\, e^{st} F(s) \quad \text{[Inverse]} \tag{3-41}$$

where s = the complex variable (or frequency) $\sigma + j\omega$.

The Fourier transform is applicable to functions that exist for all time t, whereas the Laplace transform is used for functions that are zero for $t < 0$. In the expression (3-41) for the inverse Laplace transform, the constant c is a convergence factor that enables one to apply the Laplace transform to functions whose Fourier transforms do not exist.

The conditions for the existence of the Fourier transform of a function, known as Dirichlet's conditions, are

(a) $f(t)$ has a finite number of discontinuities in a finite interval

(b) $f(t)$ has a finite number of infinite-valued points in a finite interval

(c) $f(t)$ has a finite number of maxima and minima in a finite interval

(d) $\int_{-\infty}^{+\infty} |f(t)|\, dt$ is finite

The conditions for the existence of the Laplace transform of a function are identical with those for the Fourier transform except that the fourth condition is relaxed to

$$\int_{-\infty}^{+\infty} |f(t)|\, e^{-ct}\, dt \text{ is finite for a finite } c$$

3-4.2 THEOREMS

The following theorems are useful for applying the Laplace and Fourier transforms to the solution of differential equations:

(a) $$\mathcal{L}[a f(t)] = a F(s) \tag{3-42}$$

(b) $$\mathcal{L}[f_1(t) \pm f_2(t)] = F_1(s) \pm F_2(s) \tag{3-43}$$

(c) $$\mathcal{L}\left[\frac{d^n f(t)}{dt^n}\right] = s^n F(s) - s^{n-1} f(0+) - s^{n-2} f^1(0+) - \ldots - s f^{(n-2)}(0+) - f^{(n-1)}(0+) \tag{3-44}$$

(d) $$\mathcal{L}\left[\int_{-\infty}^{t} \overset{(n \text{ times})}{\cdots} \int_{-\infty}^{t} f(t)\, dt\right] = \frac{F(s)}{s^n} + \frac{\int_{-\infty}^{0+} f(t)\, dt}{s^n} + \frac{\int_{-\infty}^{0+} \int_{-\infty}^{t} f(t)\, dt}{s^{n-1}} + \ldots + \frac{\int_{-\infty}^{0+} \left[\int_{-\infty}^{t} \overset{(n-1 \text{ times})}{\cdots} \int_{-\infty}^{t} f(t)\, dt\right]}{s} \tag{3-45}$$

(e) $$\mathcal{L}\left[f\left(\frac{t}{a}\right)\right] = a F(as) \tag{3-46}$$

(f) $$\mathcal{L}\left[\int_{0}^{t} f_1(t-\tau) f_2(\tau)\, d\tau\right] = F_1(s) F_2(s) \tag{3-47}$$

(g) $$\mathcal{L}[f(t-a)] = e^{-as} F(s) \quad \text{if } f(t-a) = 0 \text{ for } 0 < t < a \tag{3-48}$$

(h) $\mathcal{L}[f(t+a)] = e^{as}F(s)$ iff $f(t+a) = 0$ for $-a < t < 0$ (3-49)

(i) $\lim_{s\to 0} s\mathcal{L}[f(t)] = \lim_{t\to\infty} f(t)$ (3-50)

(j) $\lim_{s\to\infty} s\mathcal{L}[f(t)] = \lim_{t\to 0} f(t)$ (3-51)

(k) $\mathcal{L}[f_1(t)f_2(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s-w)F_2(w)dw$ (3-52)

(l) $\mathcal{L}[f_1(t)f_2(t)] \neq F_1(s)F_2(s)$ (3-53)

Theorems (a), (b), (e), (f), (i), (k), and (l) also apply to the Fourier transform. Theorem (c) is called the *red differentiation* theorem and theorem (d) is called the *real integration* theorem. Theorem (e) is used to change the time scale of a problem and is called the *normalization* theorem. Theorem (f) is the *real convolution* theorem and, if applied to the convolution integral [Eq. (3-35)], yields the very important result

$$\mathcal{F}[x(t)] = X(s) = \mathcal{F}\left[\int_{-\infty}^{t} dt_1\, w(t-t_1)\, y(t_1)\right] = W(s)Y(s) \quad (3\text{-}54)$$

Theorems (g) and (h) apply to the Fourier transform without the stated restrictions. Theorem (i) is called the *final-value* theorem and theorem (j) is called the *initial-value* theorem. Theorems (k) and (l) are included primarily to prevent the common error suggested by theorem (l), namely, *incorrectly* stating that the transform of the product of two time functions is the product of the separate transforms of the functions.

3-4.3 SOLUTION OF DIFFERENTIAL EQUATIONS

The solution of ordinary linear differential equations is accomplished by means of theorems (a), (b), (c), and (d). Applying these theorems to Eq. (3-1), one obtains

$$\left[\sum_{i=0}^{n} a_i s^i\right] X(s) - A(s) = \left[\sum_{j=0}^{m} b_j s^j\right] Y(s) + B(s) \quad (3\text{-}55)$$

where $A(s)$ is a polynomial in s depending upon the a's and the initial values of x and its first $(n-1)$ derivatives, and $B(s)$ is a polynomial in s depending upon the b's and the initial values of y and its first $(m-1)$ derivatives. The response transform can be obtained by solving Eq. (3-55) for $X(s)$

$$X(s) = \left[\frac{\sum_{j=0}^{m} b_j s^j}{\sum_{i=0}^{n} a_i s^i}\right] Y(s) + \left[\frac{B(s) + A(s)}{\sum_{i=0}^{n} a_i s^i}\right] \quad (3\text{-}56)$$

In words, this equation can be written

$$\begin{pmatrix}\text{response}\\ \text{(transform)}\end{pmatrix} = \begin{pmatrix}\text{system}\\ \text{(function)}\end{pmatrix}\begin{pmatrix}\text{input}\\ \text{transform}\end{pmatrix} + \begin{pmatrix}\text{initial condition}\\ \text{function}\end{pmatrix} \quad (3\text{-}57)$$

The ratio of the response transform to the input transform when all initial conditions are zero (i.e., the initial condition function is zero) is called the *system function* or the *transfer function* of the system. This function depends only upon the coefficients of the differential equation and is independent of the input and the initial conditions. Comparing Eq. (3-57) (with initial condition function set equal to zero) with Eq. (3-54), it is evident that the transfer function of a system equals the transform of the impulse response of the system for a unit impulse.

Transforming a differential equation enables the analyst to replace the processes of differentiation and integration by simple algebraic processes. Then, the transform $X(s)$ can be found algebraically. Subsequently, the system response $x(t)$ corresponding to the response transform $X(s)$ can be found by using the inversion theorem [Eq. (3-41)]. However, this theorem usually involves contour integration in the complex s plane. To avoid this integration, tables of transform pairs have been constructed that give the time function corresponding to a given transform directly. A brief list of commonly used transform pairs is given in Table 3-1. More extensive tables can be found in references (13), (20), and (21).

TABLE 3-1 LAPLACE TRANSFORM PAIRS

$F(s)$	$f(t), t \geqq 0$
1	$\delta_0(t)$, unit impulse
$\frac{1}{s}$	$\delta_{-1}(t)$, unit step
$\frac{1}{s^2}$	$\delta_{-2}(t)$, unit ramp
$\frac{1}{Ts+1}$	$\frac{1}{T} e^{-t/T}$
$\frac{\omega}{s^2+\omega^2}$	$\sin \omega t$
$\frac{s}{s^2+\omega^2}$	$\cos \omega t$
$\frac{1}{s^2+2\zeta\omega_n s+\omega_n^2}$	(1) $\zeta < 1$: $\frac{1}{\omega_n\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1-\zeta^2}t$ (2) $\zeta = 1$: $te^{-\omega_n t}$ (3) $\zeta > 1$: $\frac{1}{\omega_n\sqrt{\zeta^2-1}} e^{-\zeta\omega_n t} \sin t\, \omega_n \sqrt{\zeta^2-1}t$
$\frac{1}{(s+\alpha)^2+\beta^2}$	$\frac{1}{\beta} e^{-\alpha t} \sin \beta t$
$\frac{s+\alpha}{(s+\alpha)^2+\beta^2}$	$e^{-\alpha t} \cos \beta t$
$\frac{1}{s^n}$	$\frac{1}{(n-1)!} t^{n-1}$
$\frac{1}{(Ts+1)^n}$	$\frac{1}{(n-1)!} \frac{t^{n-1}}{T^n} e^{-t/T}$

If tables of transform pairs are unavailable, or if the particular transform whose inverse is sought is not listed in the tables, the method of partial fractions may be used to expand the transform into a sum of terms, each of which is readily recognized as the transform of a simple time function. If the transform whose inverse is sought is a ratio of rational polynomials, the roots of the numerator polynomial are called the ***zeros*** of the function and the roots of the denominator polynomial are called the ***poles*** of the function. If the poles of the function are not repeated, they are called ***single-order*** poles. The order of a ***multiple-order*** pole is the number of times the pole is repeated. For a function containing only single-order poles, the partial-fraction expansion of the function is

$$F(s) \triangleq \frac{N(s)}{D(s)} = \sum_{k=1}^{n} \frac{K_k}{s - s_k} \tag{3-58}$$

where

$$K_k \triangleq \left[\frac{(s - s_k)N(s)}{D(s)}\right]_{s=s_k} = \left[\frac{N(s)}{D'(s)}\right]_{s=s_k} \tag{3-59}$$

and s^k is the kth root of the denominator polynomial $D(s)$.

If the transform contains multiple-order poles, the partial-fraction expansion of the function is

$$F(s) \triangleq \frac{N(s)}{D(s)} = \sum_{k=1}^{n} \sum_{j=1}^{m_k} \frac{K_{kj}}{(s - s_k)^{m_k - j + 1}} \tag{3-60}$$

where

$$K_{kj} \triangleq \frac{1}{(j-1)!} \left\{ \frac{d^{j-1}}{ds^{j-1}} \left[\frac{(s - s_k)^{m_k} N(s)}{D(s)} \right] \right\}_{s=s_k} \tag{3-61}$$

and m_k is the order of the pole of $F(s)$ at $s = s_k$.

From Eqs. (3-58) and (3-60), it is obvious that the expansion of a rational function when inverted produces a sum of exponential terms for the corresponding time function. Terms containing exponentials with complex arguments will appear in conjugate pairs and can therefore be combined to form product terms (exponential multiplied by a sine or cosine function) representing damped sinusoids.

Example. The system defined by the equation

$$\frac{d^4x}{dt^4} + 10.65\frac{d^3x}{dt^3} + 89.0\frac{d^2x}{dt^2} + 15.50\frac{dx}{dt} + 27.0x = 27.0y \tag{3-62}$$

is initially at rest. At $t = 0$, a unit ramp input is applied. Find the difference between the input y and the output x as a function of time.

Solution. Since the system is initially at rest, all initial conditions are zero. Transforming Eq. (3-62) results in

$$X(s) = \frac{27.0}{s^4 + 10.65s^3 + 89.0s^2 + 15.50s + 27.0} Y(s) \tag{3-63}$$

Let

$$e(t) = y(t) - x(t) \tag{3-64}$$

Then, transforming Eq. (3-64) and substituting for X(s) from Eq. (3-63), E(s) becomes

$$E(s) = \frac{s[s^3 + 10.65s^2 + 89.0s + 15.501}{s^4 + 10.65s^3 + 89.0s^2 + 15.50s + 27.0} Y(s) \tag{3-65}$$

By referring to the values of the roots given in Eq. (3-14), the denominator $D(s)$ of Eq. (3-65) can be factored as follows:

$$D(s) = (s + 0.0702 - j0.552)(s + 0.0702 + j0.552)$$
$$(s + 5.26 - j7.72)(s + 5.26 + j7.72) \tag{3-66}$$

The transform of $y(t)$, found from Table 3-1, is

$$Y(s) = \frac{1}{s^2} \tag{3-67}$$

Using the factored form of the denominator and substituting $1/s^2$ for the value of $Y(s)$ in Eq. (3-65) results in

$$E(s) = \frac{[s^3 + 10.659 + 89.0s + 15.501}{s[s + 0.0702 - j0.552]\,[s + 0.0702 + j0.552]\,[s + 5.26 - j7.72]\,[s + 5.26 + j7.72]} \tag{3-68}$$

Since two pairs of the poles of $E(s)$ appear as conjugate pairs, the partial-fraction expansion of $E(s)$ can be written

$$E(s) = \frac{K_1}{s} + \frac{K_2}{s + 0.0702 - j0.552} + \frac{\overline{K_2}}{s + 0.0702 + j0.552} + \frac{K_3}{s + 5.26 - j7.72} + \frac{\overline{K_3}}{s + 5.26 + j7.72} \tag{3-69}$$

where a bar over a constant indicates the complex conjugate of the constant. Using the expansion theorem [Eq. (3-59)]

$$K_1 = \frac{15.50}{27.0} = 0.574 \tag{3-70}$$

$$K_2 = \left[\frac{(s^3 + 10.65s^2 + 89.0s + 15.50)}{s\,(s + 0.0702 + j0.552)\,(s^2 + 10.51s + 87.2)}\right]_{s=-0.0702+j0.552} = 0.918\, e^{-j0.602\pi} \tag{3-71}$$

$$\overline{K_2} = 0.918\, e^{+j0.602\pi} \tag{3-72}$$

$$K_3 = \left[\frac{(s^3 + 10.65s^2 + 89.0s + 15.50)}{s\,(s^2 + 0.1403s + 0.310)\,(s + 5.26 + j7.72)}\right]_{s=-5.26+j7.72} = 2.89 \times 10^{-4}\, e^{-j0.277\pi} \tag{3-73}$$

$$\overline{K_3} = 2.89 \times 10^{-4} e^{+j0.277\pi} \tag{3-74}$$

Inverse transforming Eq. (3-69)

$$e(t) = 0.574 + 0.918\, e^{-0.0702t} [e^{j(0.552t - 1.89)} + e^{-j(0.552t - 1.89)}] + 2.89 \times 10^{-4}\, e^{-5.26t} [e^{j(7.72t - 0.87)} + e^{-j(7.72t - 0.87)}] \quad (3\text{-}75)$$

The bracketed functions on the right side of Eq. (3-75) are recognized as cosine functions, so that $e(t)$ can be written as

$$e(t) = 0.574 + 1.836\, e^{-0.0702t} \cos (0.5525 - 1.89) + 5.78 \times 10^{-4}\, e^{-5.26t} \cos (7.72t - 0.87) \quad (3\text{-}76)$$

It is convenient for plotting purposes to write the arguments of the cosine functions in degrees and to use trigonometric identities to reduce the phase angles to angles smaller than 45°. If this is done, $e(t)$ can be written as

$$e(t) = 0.574 + 1.836\, e^{-0.0702t} \sin (31.6t - 18.3)^\circ - 5.78 \times 10^{-4}\, e^{-5.26t} \sin (442t + 40.2)^\circ \quad (3\text{-}77)$$

3-4.4 FREQUENCY RESPONSE

It is often important to find the response of a system to a sinusoidal input. For a sinusoidal input, the output of the system will also be sinusoidal after transients have died out. The amplitude and phase angle of the output relative to the input are dependent only upon $W(s)$, the transfer function of the system, and can be determined by letting $s = j\omega$ in the transfer function, where ω is the frequency (in radians/second) of the input sinusoid. The ratio of output amplitude to input amplitude is then given by

$$\frac{A_x}{A_y} = |W(j\omega)| \quad (3\text{-}78)$$

where A_x is the output amplitude, A, is the input amplitude, and $W(s)$ is the transfer function of the system. The phase angle of the output relative to the phase angle of the input is given by†

$$\phi_x - \phi_y = \angle\, W(j\omega) \quad (3\text{-}79)$$

where ϕ_x is the output phase angle and ϕ_y is the input phase angle.

When the transfer function of a system is evaluated as a function of frequency for a sinusoidal input, the complex function that results is called the ***frequency response*** of the system.

†Symbol $\angle$ denotes "angle"

3-5 BLOCK DIAGRAMS AND SIGNAL-FLOW GRAPHS [22,23,24,25,26,27,28]

3-5.1 BLOCK DIAGRAMS

Equations (3-54) and (3-57) demonstrate that, with zero initial conditions, the transform of the output of a system can be expressed in terms of the input transform and the system function. The system function can be thought of as an ***operator.*** That is, the system function operates on the input transform to produce the output transform. In a similar manner, the system operates on the input to produce the output in the time domain, the operation being defined by the convolution integral [Eq. (3-31)] and depending only upon the impulse response of the system. The concept of an operator is presented pictorially by the technique known as ***operational block diagram algebra.*** The block diagram of a system is the pictorial representation of the mathematical operations involved in the differential equations that describe the system.

Table **3-2** presents a list of symbols used in the block diagram representation of a system and Fig. **3-6** presents a list of reductions that enable one to simplify or reduce the block diagrams of a system. Since the block diagram contains no more information than the differential equations, the manipulation of a block diagram is merely a pictorial process of manipulating the differential equations. The advantage of a block diagram representation is that the operational relations in a system are emphasized rather than the hardware. By becoming familiar with common block arrangements, the designer can interpret the function of various elements in a system much more rapidly than would be possible from an inspection of the differential equations.

Example. The transformed equations of a servomotor driving an inertia load coupled to the motor through a flexible shaft are

$$T_m = (J_m s^2 + f_m s)\,\theta_m + K(\theta_m - \theta_L) \quad (3\text{-}80)$$

$$K(\theta_m - \theta_L) = J_L s^2 \theta_L + T_L \quad (3\text{-}81)$$

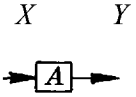
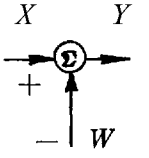
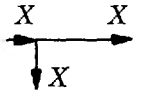
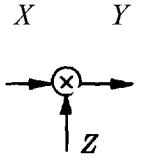
where

T_m = motor torque
K = shaft stiffness
J_m = motor inertia
θ_L = load angle
f_m = motor damping
J_L = load inertia
θ_m = motor angle
T_L = load torque

The damping of the flexible shaft is assumed to be negligible. Draw the block diagram of the system and reduce the diagram, keeping the motor angle θ_m and the load angle θ_L in evidence.

Solution. The block diagram of the system is drawn in its "primitive" form in Fig. **3-7A.** The successive steps necessary to reduce the "primitive" diagram to the desired form are shown in Figs. **3-7B** to **3-71** with the rules used for each step indicated below each step.

TABLE 3-2 BLOCK DIAGRAM SYMBOLS

Symbol	Description	Operation
	variable	
	operator	$Y = AX$
	summing point	$Y = X - W$
	splitting point	$X = X$
	multiplier	$Y = XZ$

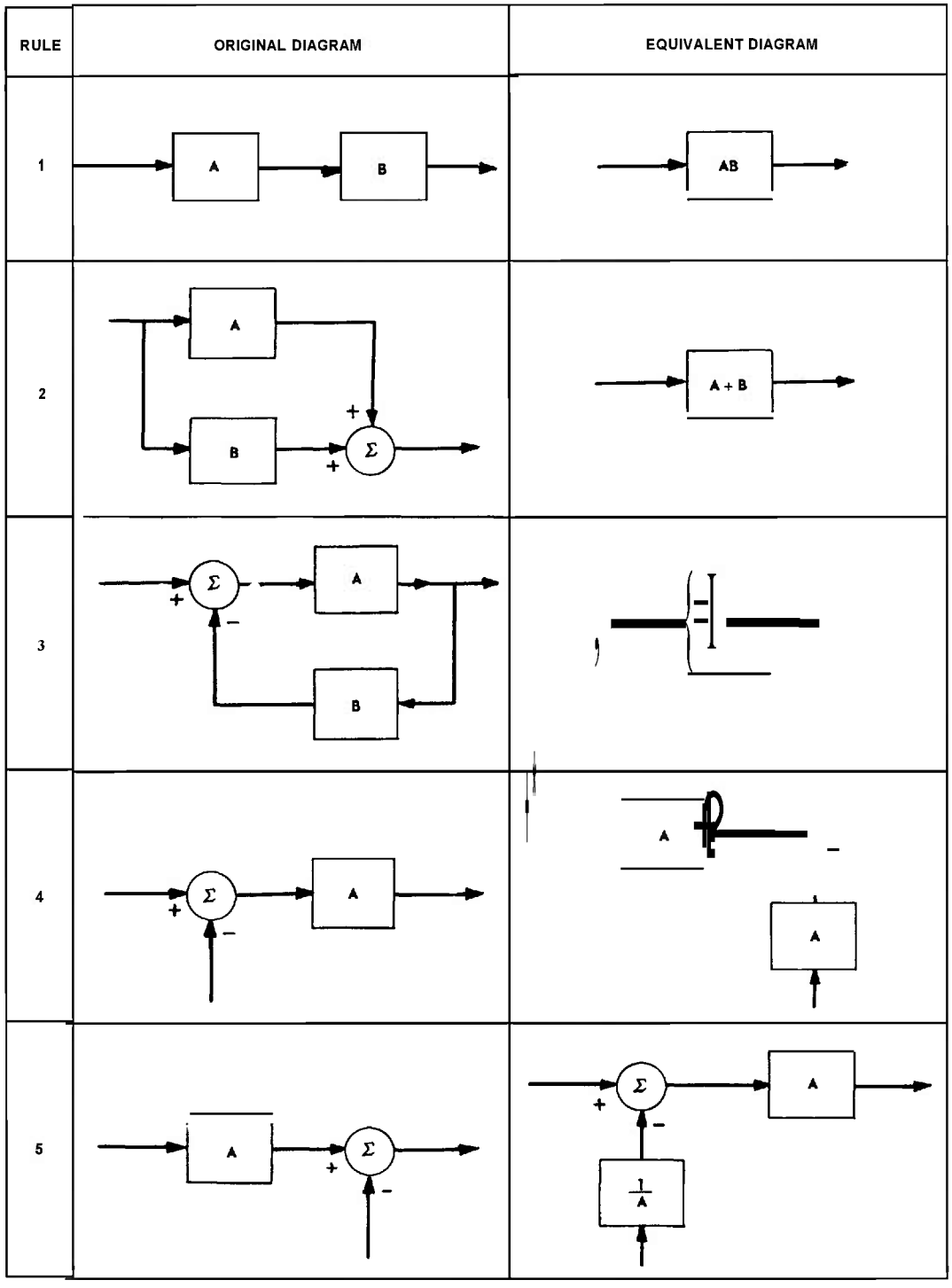


Fig. 3-6 Block diagram manipulation and reduction "rules". (Sheet 1 of 3)

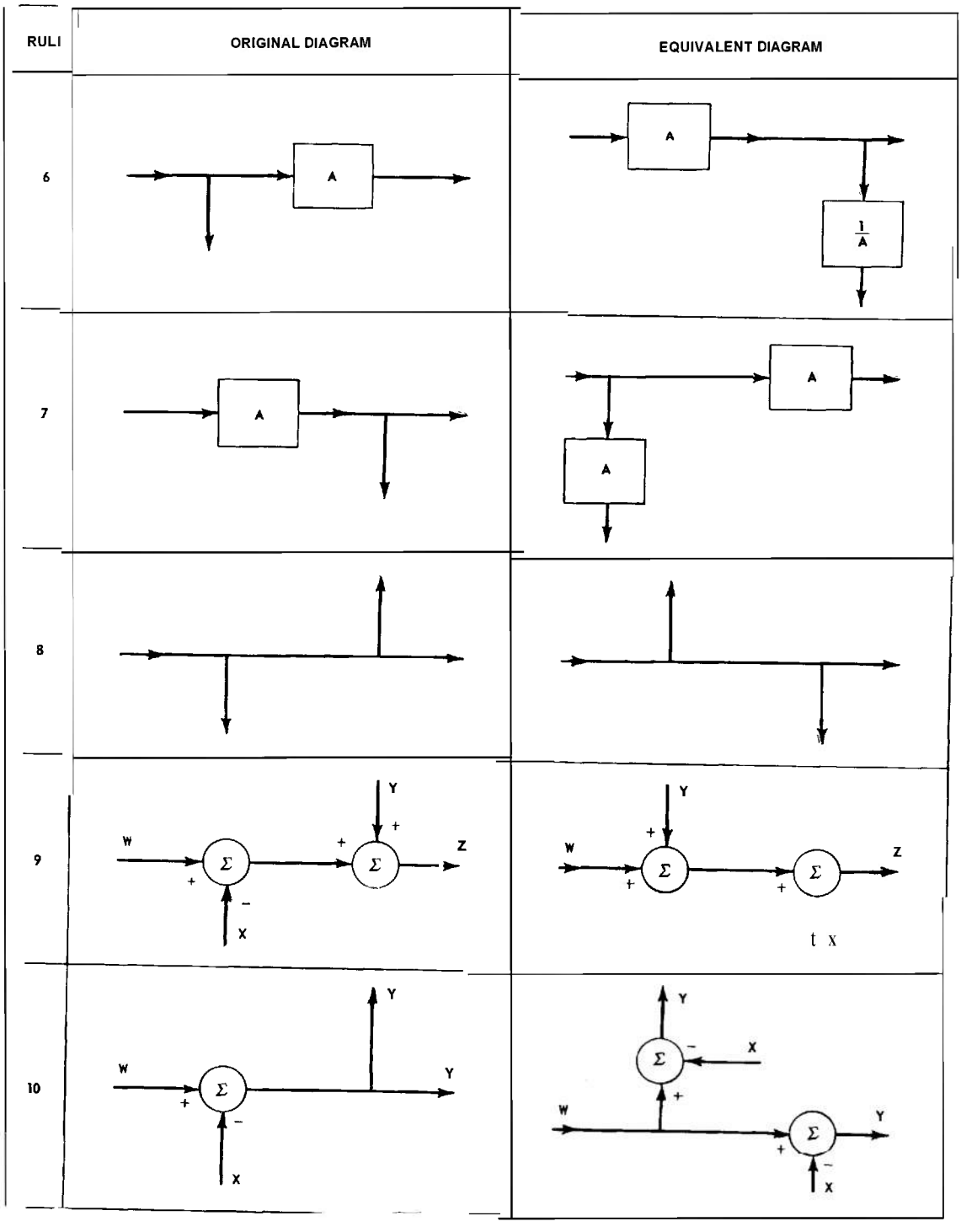


Fig. 3-6 Block diagram manipulation and reduction "rules". (Sheet 2 of 3)

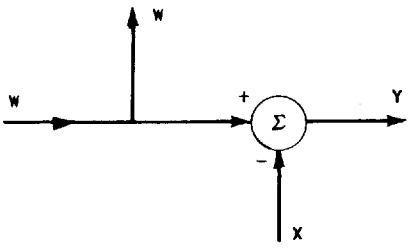
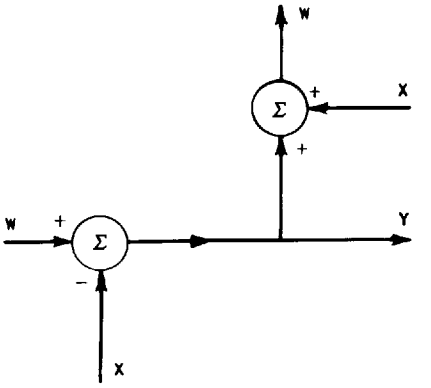
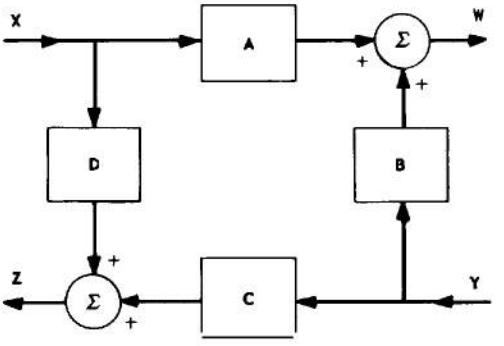
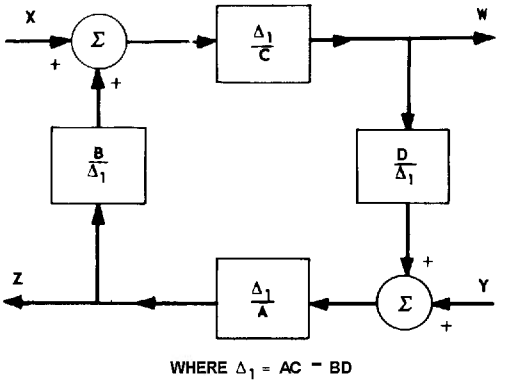
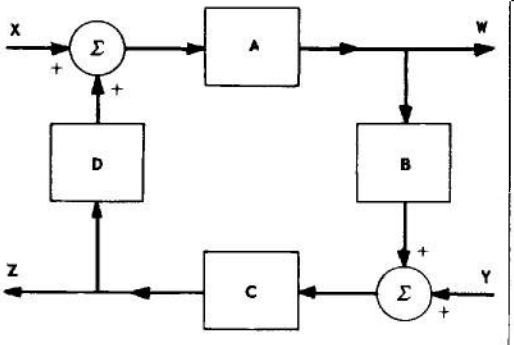
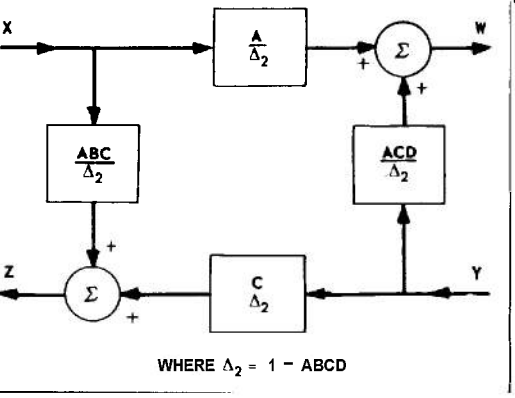
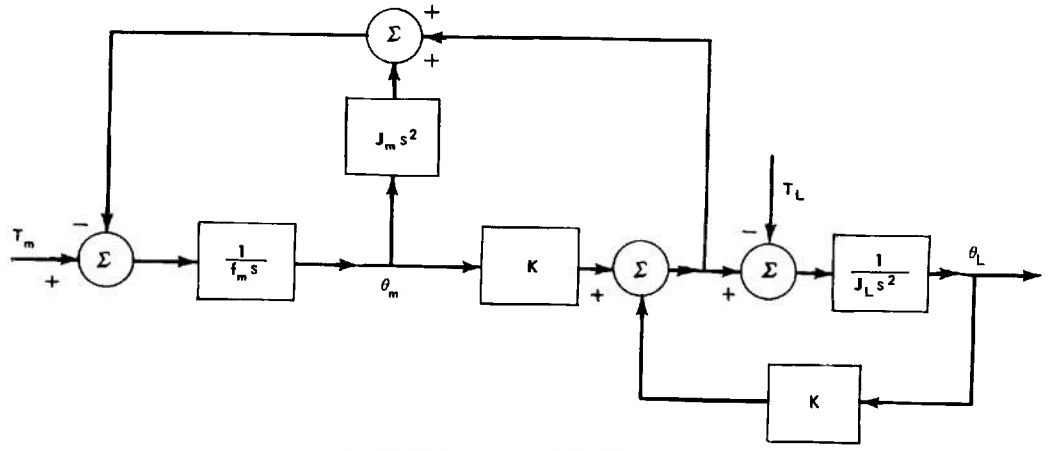
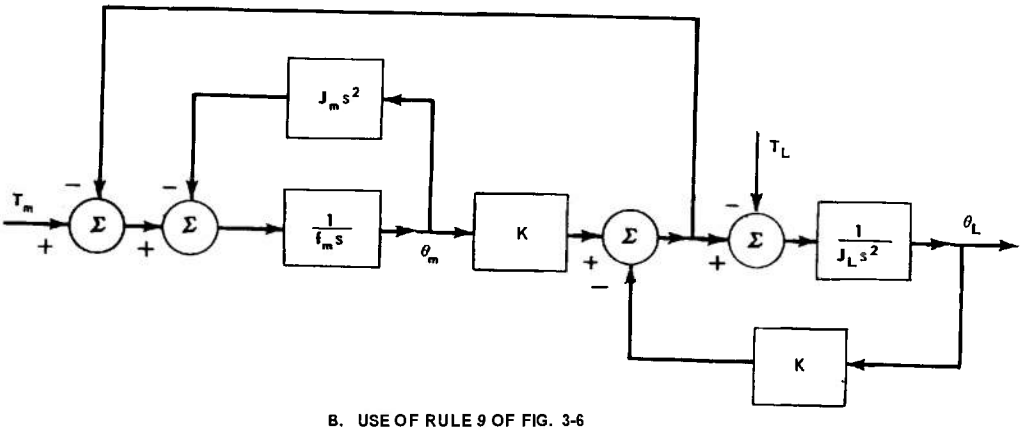
RULE	ORIGINAL DIAGRAM	EQUIVALENT DIAGRAM
11		
12		
13		

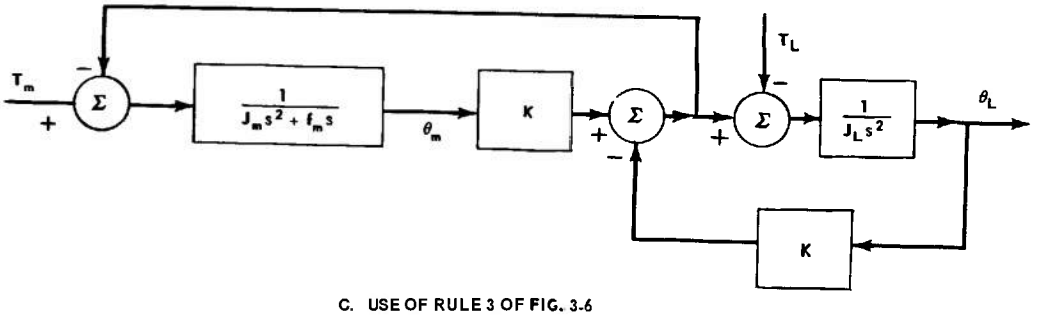
Fig. 3-6 Block diagram manipulation and reduction "rules". (Sheet 3 of 3)



A. ELEMENTARY BLOCK DIAGRAM

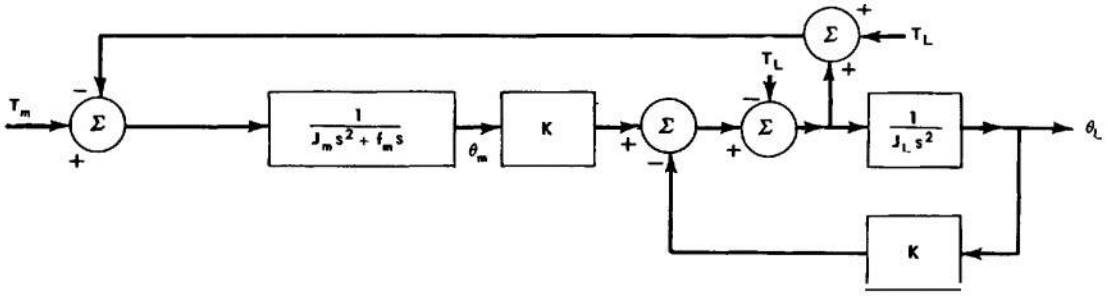


B. USE OF RULE 9 OF FIG. 3-6



C. USE OF RULE 3 OF FIG. 3-6

Fig. 3-7 Block diagram examples. (Sheet 1 of 3)



D. USE OF RULE 11 OF FIG 3-6

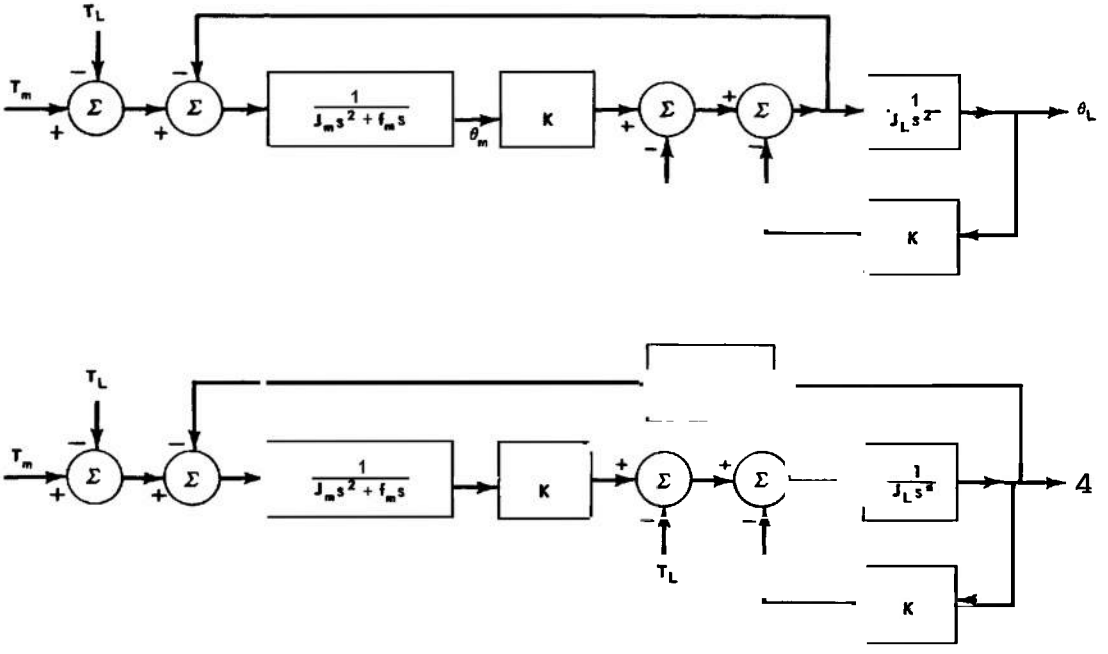
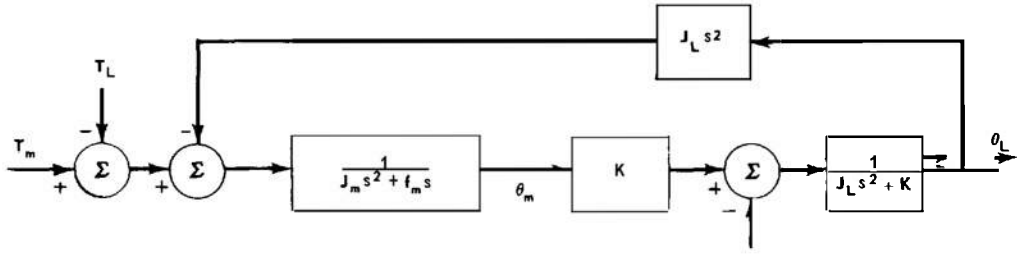
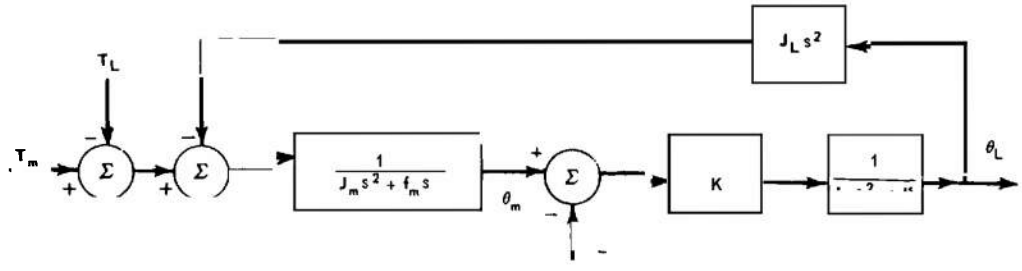
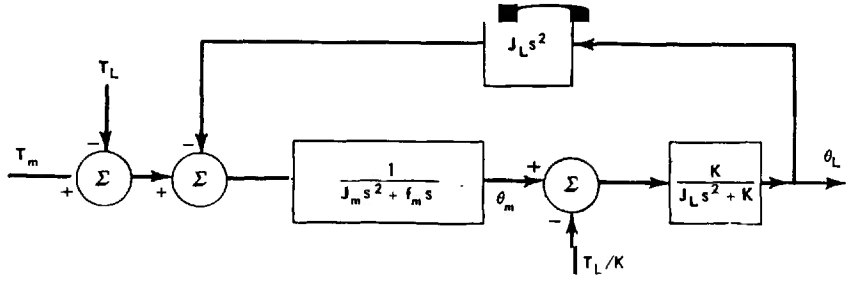


Fig. 3-7 Block diagram examples. (Sheet 2 of 3)





H. USE OF RULE OF FIG. 3-6



I. USE OF RULE 1 OF FIG. 3-6

fig. 3-7 Block diagram examples. (Sheet 3 of 3)

3-5.2 SIGNAL-FLOW GRAPHS[27,28]

An alternate procedure for representing the differential equations of a system pictorially is Mason's[27] signal-flow graph method. In a signal-flow graph, variables are represented by points called *nodes* and transfer functions are represented by directed lines or branches called *transmittances*. The distinction between the summing points and the splitting points of block diagram algebra is eliminated in the signal-flow graph. The rules for drawing a signal-flow graph are as follows[28]:

(a) Signals travel along branches only in the direction of the arrows.

(b) A signal traveling along any branch is multiplied by the transmittance of that branch.

(c) The value of the variable represented by any node is the sum of all signals entering the node.

(d) The value of the variable represented by any node is transmitted on all branches leaving that node.

Example. The two equations

$$x_1 = t_{01}\,x_0 + t_{11}\,x_1 + t_{21}\,x_2 \qquad \textbf{(3-82)}$$

$$x_2 = t_{02}\,x_0 + t_{12}\,x_1 + t_{22}\,x_2 \qquad \textbf{(3-83)}$$

are represented by a signal-flow graph in Fig. **3-8.**

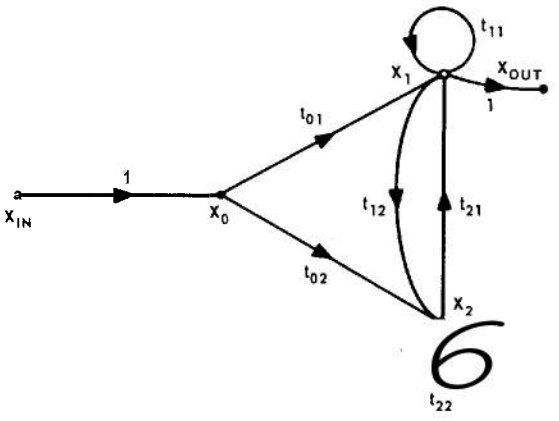


Fig. 3-8 Signal-flow graph in three variables.

For convenience, the signal-flow graph is usually drawn such that no branch enters an input node or leaves an output node. This is accomplished by introducing an additional node connected by a unity-transmittance branch to each input and output node as shown in Fig. **3-8,** where the input node is assumed to be x_0 and the output node is assumed to be x_1.

The *order* of a signal-flow graph is a measure of the number of independent feedback loops and thus indicates the complexity of the system. The order of the signal-flow graph is the minimum number of *essential nodes* — those nodes that must be removed to eliminate all feedback paths. A node is removed either by setting the variable associated with the node equal to zero or by deleting all branches leaving the node. Signal-flow graphs of orders one and two are shown in Figs. 3-9 and **3-10,** respectively. The signal-flow graph of Fig. **3-8** is of order two, the essential nodes being x_1 and x_2.

The reduction of signal-flow graphs is accomplished by application of the following rules[28]:

(a) Two parallel paths may be replaced by a single path with a transmittance equal to the sum of the two original transmittances (Fig. **3-11).**

(b) Two cascaded paths are equivalent to a single path with a transmittance equal to the product of the two original transmittances (Fig. **3-12).**

(c) The *termination* of a branch with transmittance t can be shifted one node forward by the following steps (Fig. **3-13)**:

(1) Determine all the branches leaving the original terminating node x of branch t.

(2) Draw new branches from the starting node x_0 of branch t to the terminating nodes of all the branches leaving the terminating node x.

(3) To each of the new branches thus drawn assign a transmittance equal to the product of t times the transmittance from node x to the node on which the new branch terminates.

(4) Eliminate the original branch t.

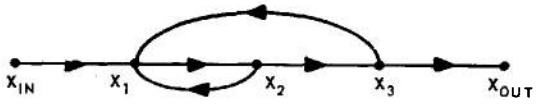
(5) Change the variable of the original node x to $x' = x - tx_0$.

(d) The ***starting point*** or ***origin*** of a branch with transmittance t can be shifted one node backward by the following steps (see Fig. **3-14**) :

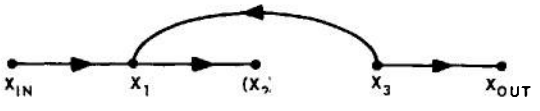
(1) Determine all the branches entering the original starting node x of branch t.

(2) Draw new branches from the starting nodes of all the branches entering starting node x to the terminating node x_i of branch t.

(3) To each **of** the new branches thus drawn assign a transmittance equal to the product of t times the transmittance from the node at which the new branch starts to node x.

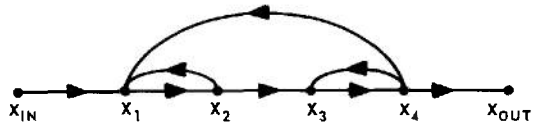


A. ORIGINAL GRAPH

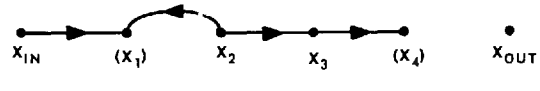


B. ESSENTIAL NODE REMOVED

Fig. 3-9 Signal-flow graph of order one.



A. ORIGINAL GRAPH



B. ESSENTIAL NODES REMOVED

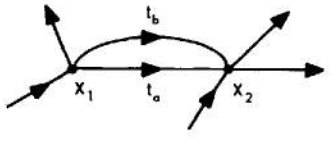
Fig. 3-70 Signal-flow graph of order two.

(4) Eliminate the original branch t.

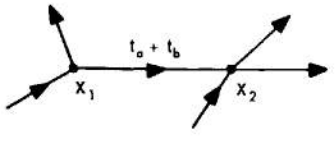
(e) A ***self-loop*** with transmittance ***t*** of a node x can be removed by dividing the transmittances of all branches entering node x by $(1 - t)$ and eliminating the loop (Fig. **3-15** ; in this figure. $t \equiv t_{22}$, where the first subscript denotes the node on which the branch originates and the second subscript denotes the node on which the branch terminates).

Note, in rule (c), that a self-loop is created at node x_0 for a branch starting from the terminating node x of branch t and ending on the starting node x_0 of branch t (Fig. **3-13** does not happen to have such a branch). In rule (d), a self-loop is created at node x_i for a branch starting from the terminating node x_i of branch t and ending on the starting node x of branch t.

Example. The various steps involved in reducing the second-order signal-flow graph of Fig. **3-8** are shown in Fig. **3-16.**

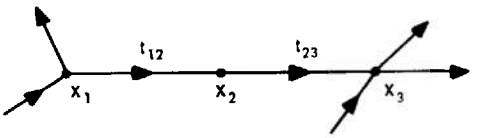


A. ORIGINAL GRAPH

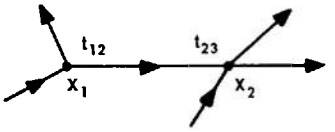


B. EQUIVALENT GRAPH

Fig. 3-11 Signal-flow graph showing addition of parallel branches.

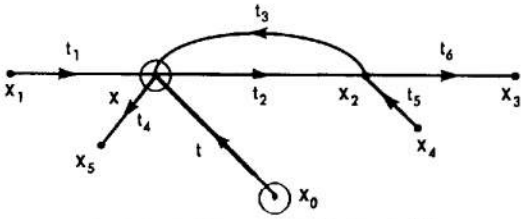


A. ORIGINAL GRAPH

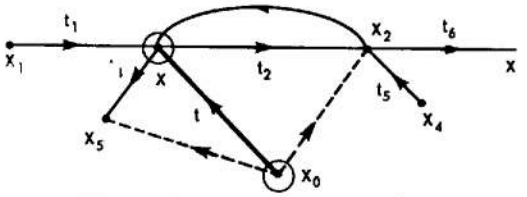


B. EQUIVALENT GRAPH

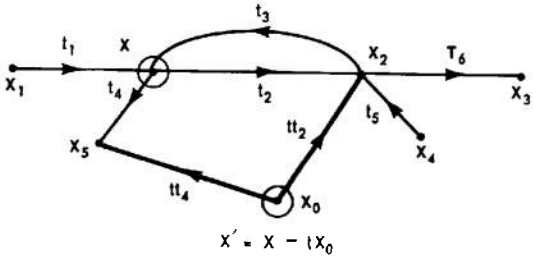
Fig. 3-12 Signal-flow graph showing multiplication of cascaded branches.



A. ORIGINAL GRAPH – t TO BE MOVED FROM X TO x_2

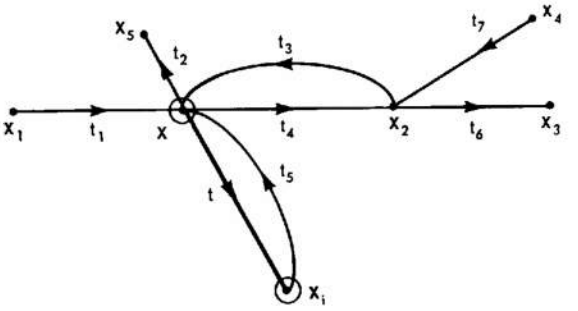


B. STEPS (1) AND (2) – INTRODUCTION OF NEW BRANCHES

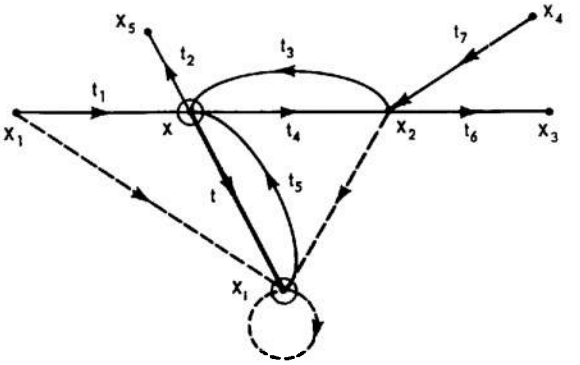


$x' = x - tx_0$

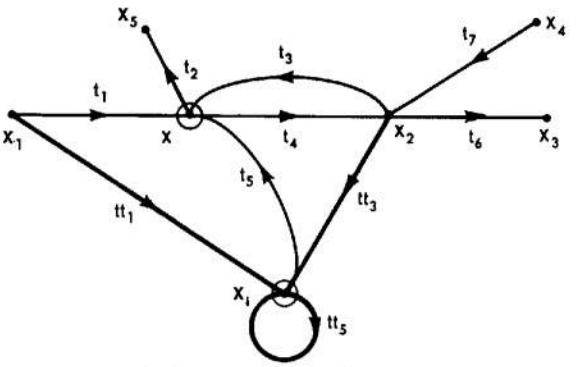
C. STEPS (3), (4), AND (5) – ELIMINATION OF OLD BRANCH; LABELLING OF NEW BRANCHES, CHANGE OF VARIABLE AT TERMINATING NODE OF OLD BRANCH



A ORIGINAL GRAPH – t TO BE MOVED FROM X TO x_1



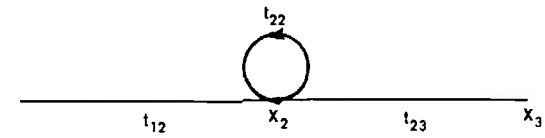
B. STEPS (1) AND (2) – INTRODUCTION OF NEW BRANCHES



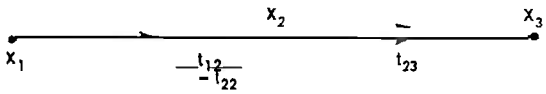
C. STEPS (3) AND (4) – ELIMINATION OF OLD BRANCH AND LABELLING OF NEW BRANCHES

Fig. 3-14 Signal-flow graph showing origin shifted one node backward.

Fig. 3-13 Signal-flow graph showing termination shifted one node forward.

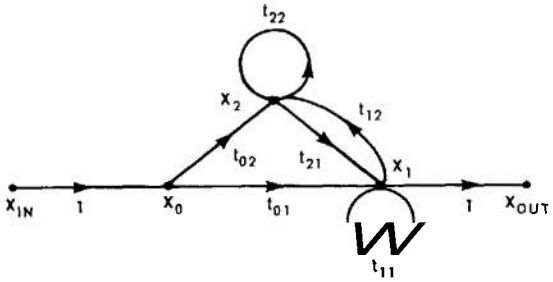


A. ORIGINAL GRAPH

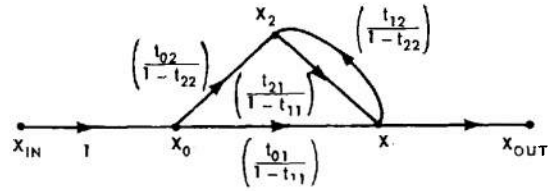


B. EQUIVALENT GRAPH

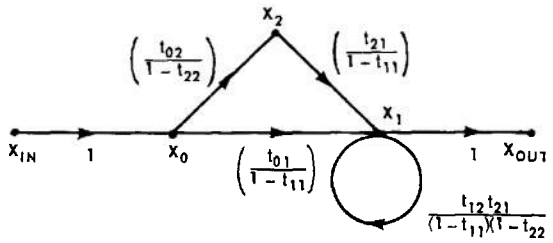
fig. *3-15* Signal-flow graph showing elimination of a self-loop.



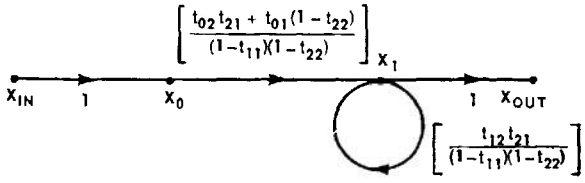
A. ORIGINAL SIGNAL-FLOW GRAPH (SECOND ORDER)



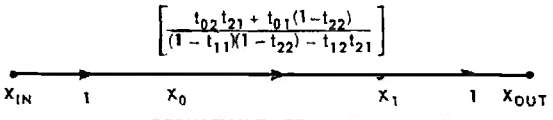
B. REDUCTION TO FIRST-ORDER GRAPH BY ELIMINATING SELF-LOOPS



C. MOVEMENT OF BRANCH $\left(\frac{t_{12}}{1-t_{22}}\right)$ TERMINATION FROM NODE X_2 TO NODE X_1



D. CASCADE AND PARALLEL BRANCHES COMBINED



E. REDUCTION TO ZERO-ORDER GRAPH BY ELIMINATION OF SELF-LOOP

fig. *3-16* Signal-flow graph showing reduction of second-order graph.

3-6 APPROXIMATE NUMERICAL AND GRAPHICAL METHODS OF DETERMINING TRANSIENT RESPONSE

A large variety of graphical and numerical procedures[29 to 45] have been proposed that enable the designer to determine the following: (1) the transient response of a system, knowing the frequency response; (2) the frequency response, knowing the transient response; or (3) the response of a system to an arbitrary input. These graphical and numerical procedures are particularly useful when purely analytical procedures are too difficult or time-consuming. When using Laplace or Fourier transforms, the difficulty that arises is primarily one of factoring high-order polynomials. When the convolution integral is involved, its direct evaluation is often time-consuming and difficult. If experimental data are available in the form of transient responses or frequency responses of system components, it is desirable to avoid the problem of approximating the data with analytical functions.

In many of the techniques, the determination of the transient response of a system from its frequency response is based on the real-part and imaginary-part integrals, i.e.,

$$f(t) = \frac{2}{\pi}\int_0^\infty d\omega \cos \omega t \; Re[F(j\omega)] \quad (3\text{-}84)$$

$$f(t) = -\frac{2}{\pi}\int_0^\infty d\omega \sin \omega t \; Im[F(j\omega)] \quad (3\text{-}85)$$

where Eq. (3-84) is the real-part integral, Eq. (3-85) is the imaginary-part integral, $f(t)$ is the time response to be evaluated, and $F(j\omega)$ is the transform of $f(t)$ evaluated for $s = j\omega$.

Floyd[41] uses Eq. (3-84) to determine the time response corresponding to a transform $F(s)$ satisfying the following conditions:

The procedure used in Floyd's method is as follows:

(a) The function $Re[F(j\omega)]$ is plotted to a linear scale and approximated by a series of straight-line segments.

(b) The straight-line approximation is written as a sum of trapezoidal functions having the general form shown in Fig. 3-17.

(c) The time response corresponding to each of these component trapezoids is given by a relation of the form

$$f_i(t) = \frac{2}{\pi} A_i \, Si(\omega_i t) \, Si(\Delta_i t) \quad (3\text{-}86)$$

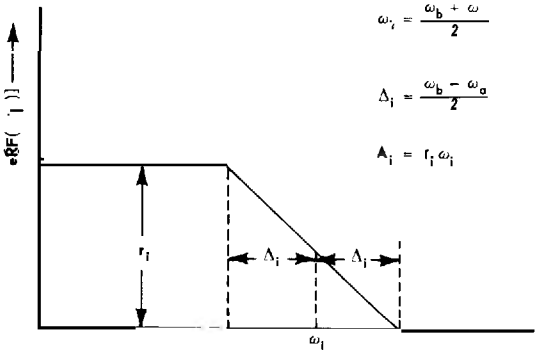
where

$$Si(x) \triangleq \frac{\sin x}{x} \quad (3\text{-}87)$$

and the quantities A_i, ω_i, and Δ_i are defined in Fig. 3-17.

(d) The total time response is obtained by taking the sum of the individual time responses corresponding to each component trapezoid, i.e., for k trapezoids

$$f(t) = \sum_{i=1}^{k} \frac{2}{\pi} A_i \, Si(\omega_i t) \, Si(\Delta_i t) \quad (3\text{-}88)$$



Example. A unity-feedback system has the closed-loop response

$$\frac{C(s)}{R(s)} = W(s) = \frac{1.4s + 0.14}{s^3 + s^2 + 1.4s + 0.14} \tag{3-89}$$

Find the impulse response $w(t)$ of the system.

Solution. $Re[W(j\omega)]$ is plotted in Fig. 3-18 and the straight-line approximation to this function is shown as the dashed curve.

Sketches of the straight-line approximation and the trapezoidal decomposition of this approximation appear in Fig. 3-19. From the approximation and the definitions of Fig. 3-17, the parameters of the trapezoids can be shown to be those listed in Table 3-3.

The impulse response is therefore given by

$$\begin{aligned} w(t) = \frac{2}{\pi} [&-0.01\, Si(0.12t)\, Si(0.12t) \\ &+ 1.71\, Si(1.06t)\, Si(0.41t) \\ &- 0.59\, Si(2.05t)\, Si(0.57t) \\ &- 1.08\, Si(4.89t)\, Si(2.272)] \end{aligned} \tag{3-90}$$

This response is plotted in Fig. 3-20 along with the exact impulse response determined by inverse transforming $W(s)$.

Guillemin[44,45] uses a different approximation to Eq. (3-84). If the real-part function is differentiated n times, the corresponding time function is found to be

$$f(t) = \frac{2}{\pi} \frac{(-1)^{\frac{n}{2}}}{t^n} \int_0^\infty d\omega \cos \omega t \, \frac{d^n Re[F(j\omega)]}{d\omega^n} \tag{3-91}$$

if n is even, and

$$f(t) = \frac{2}{\pi} \frac{(-1)^{\frac{n+1}{2}}}{t^n} \int_0^\infty d\omega \sin \omega t \, \frac{d^n Re[F(j\omega)]}{d\omega^n} \tag{3-92}$$

if n is odd.

TABLE 3-3 PARAMETERS OF TRAPEZOIDS

Trapezoid No.	ω_i	Δ_i	A_i
I	0.12	0.12	−0.01
II	1.06	0.41	+1.71
III	2.05	0.57	−0.59
IV	4.89	2.27	−1.08

If a straight-line approximation to $Re[F(j\omega)]$ is differentiated twice, the second derivative of the approximation is a series of impulses in the frequency domain extending over positive and negative frequencies. Thus

$$\frac{d^2 Re[F(j\omega)]}{d\omega^2} \cong a_0 \delta_0(\omega) + \sum_{j=1}^{m} a_j [\delta_0(\omega - \omega_j) + \delta_0(\omega + \omega_j)] \tag{3-93}$$

where the δ's are unit impulses.

The time response can then be found from

$$f(t) = -\frac{2}{\pi t^2} \left[\frac{a_0}{2} + \sum_{j=1}^{m} a_j \cos \omega_j t \right] \tag{3-94}$$

As checks on the approximations of Floyd and Guillemin, the following relations hold:

$$f(0) = \frac{2}{\pi} \int_0^\infty d\omega \, Re[F(j\omega)] = \sum_{n=1}^{k} \frac{2}{\pi} A_n \tag{3-95}$$

where A_n is the area of the nth trapezoid in Floyd's approximation.

$$\frac{a_0}{2} + \sum_{j=1}^{m} a_j = 0 \tag{3-96}$$

where a_j is the magnitude of the jth impulse in Guillemin's approximation.

$$f(0) = \frac{1}{\pi} \sum_{j=1}^{m} a_j \omega_j^2 \tag{3-97}$$

where a_j is the magnitude and ω_j is the frequency location of the jth impulse in Guillemin's approximation.

Stallard[38] has suggested a method for obtaining the time response applicable to control systems having an oscillatory response. In particular, the response $c(t)$ of a control system whose input $r(t)$ is a unit step $\delta_{-1}(t)$ can be approximated from

$$c(t) = \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{Re[W(jn\omega_1)] \sin n\omega_1 t}{n} \qquad (3\text{-}98)$$

where

$W(s) = \dfrac{C(s)}{R(s)}$, the closed-loop transfer function

$\omega_1 = \dfrac{\omega_0}{9}$

ω_0 = cutoff frequency at which the phase angle of $W(j\omega)$ is $-90°$

By using about eight terms of the series [Eq. (3-98)], an accurate representation of the step response of the system is obtained that is valid over the interval

$$0 < t < \frac{\pi}{\omega_1}$$

If the impulse response or the ramp response of the system is sought, Eq. (3-98) can be differentiated or integrated term-by-term to obtain the desired response.

When using the real-part integral [Eq. (3-84)] and the various approximations derived from it, a convenient method of finding $Re[W(j\omega)]$ is often desired when the impulse response of a feedback control system is sought. Since many of the design procedures discussed in Chs. 5 and 6 employ a graphical representation of the open-loop response $C(j\omega)/E(j\omega)$ in the gain-phase plane (see Ch. 5), Fig. 3-21 is included. This chart presents contours of constant $Re[W(j\omega)]$ on the gain-phase plane. If the open-loop frequency response $C(j\omega)/E(j\omega)$ is plotted on this chart, the intersection of the $C(j\omega)/E(j\omega)$ function with the $Re[W(j\omega)]$ contours at each frequency determines the real part of the closed-loop response $W(j\omega)$ as a function of frequency.

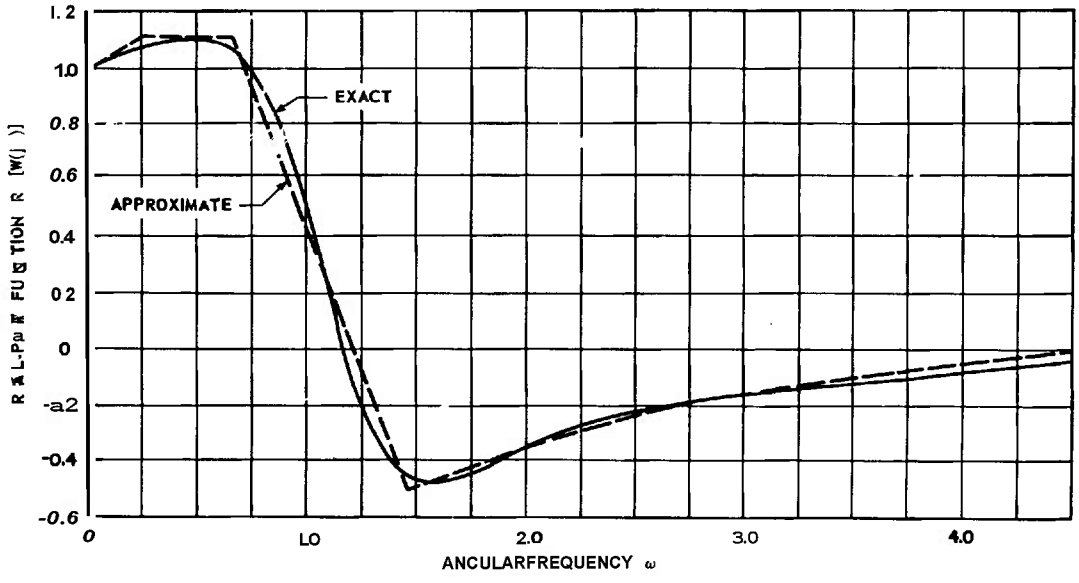
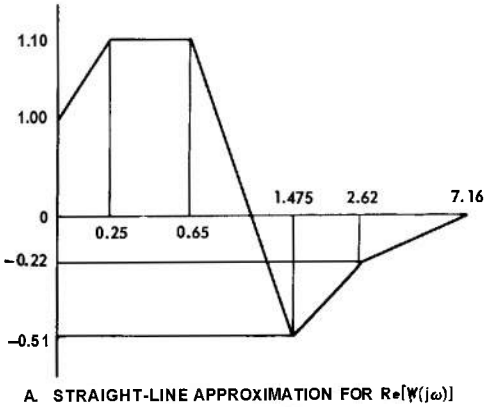
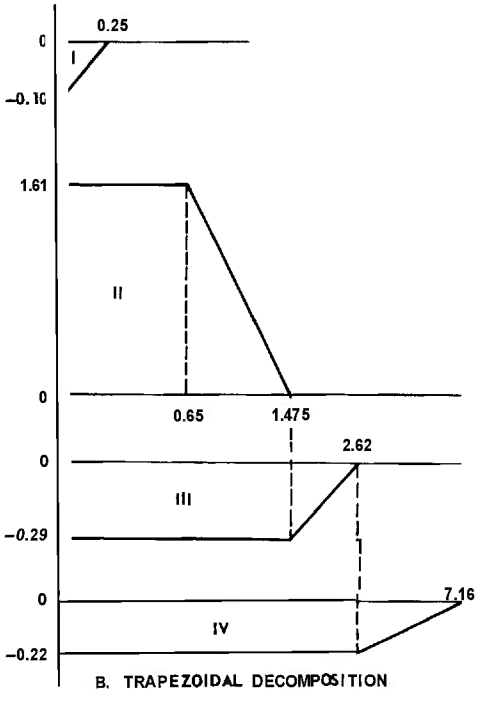


Fig. 3-78 *Real-part* function for $W(s) = \dfrac{1.4s + 0.14}{s^3 + s^2 + 1.4s + 0.14}$



A. STRAIGHT-LINE APPROXIMATION FOR Re[W(jω)] OF FIG. 3-18



B. TRAPEZOIDAL DECOMPOSITION

Fig. 3-19 Trapezoidal approximation for $Re[W(j\omega)]$.

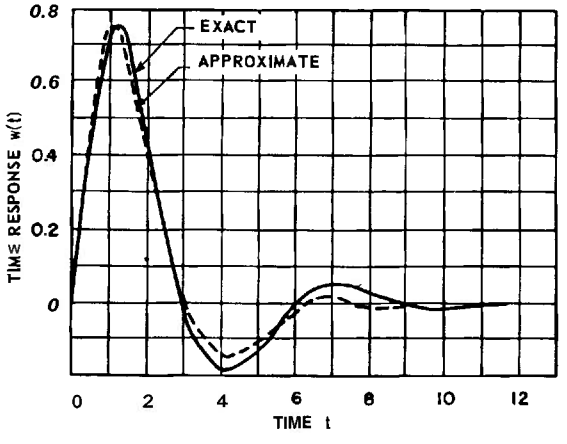


Fig. 3-20 Impulse response from Floyd's method

$$W(s) = \frac{1.4s + 0.14}{s^3 + s^2 + 1.4s + 0.14}$$

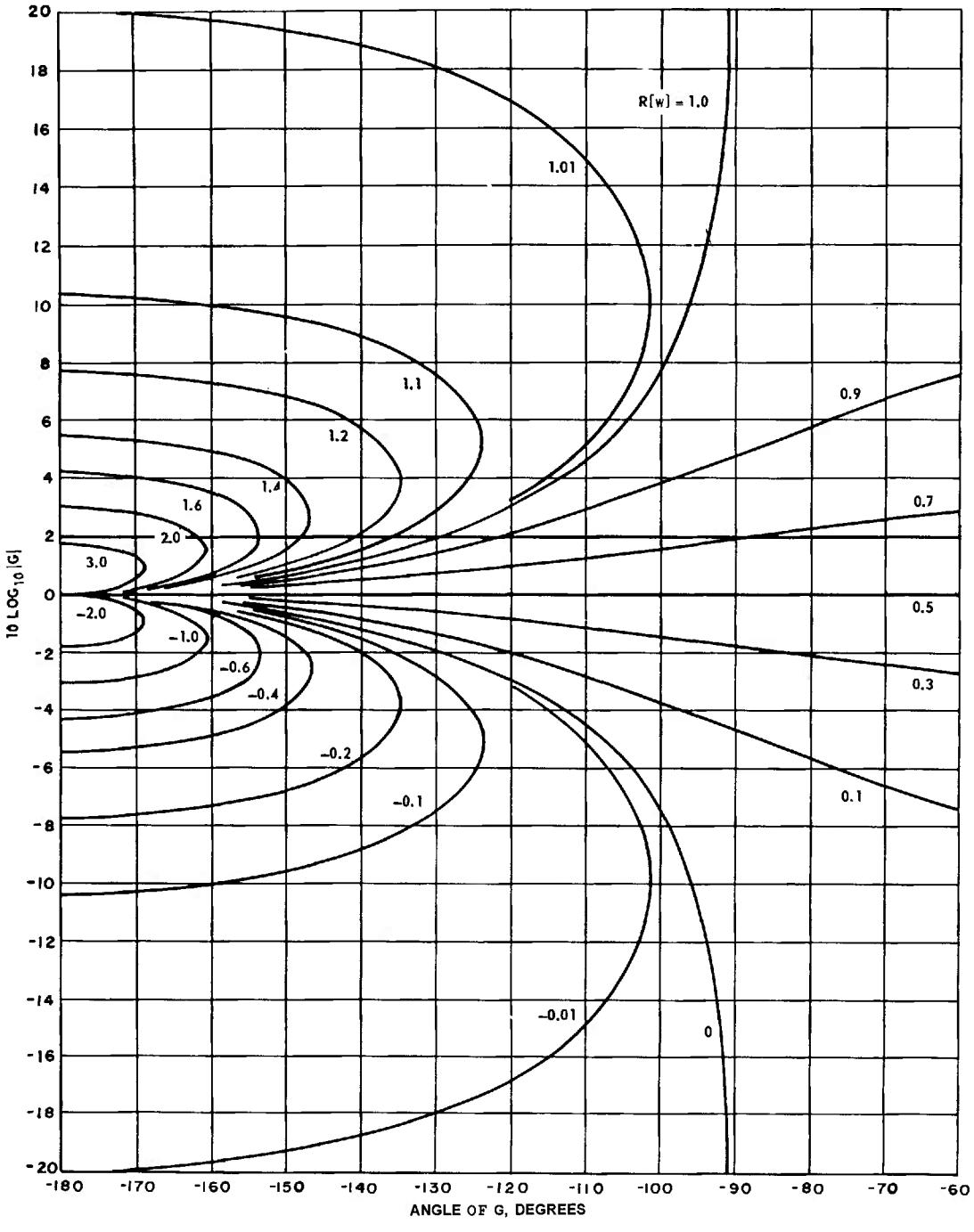


Fig. 3-27 Gain-phase loci of constant real part of $W = \dfrac{G}{1+G}$

3-7 ERROR COEFFICIENTS FOR DETERMINING RESPONSE TO AN ARBITRARY INPUT

The convolution integral [Eq. (3-31)] gives the response of a linear system to any arbitrary input. In many control applications, the input function is specified as an arbitrary function of time that cannot be classified as a transient or sinusoidal input. Since the evaluation of the convolution integral usually involves a tedious graphical or analytical procedure, methods of simplifying the calculation of the response of a unity-feedback system to an arbitrary input have been devised.

One of the most useful techniques is called the *error-coefficient method.*[40,46,47] In this procedure, the convolution integral is expanded in a Taylor series. To insure convergence of the expansion, it is assumed that the input and its derivatives have no discontinuities in the time interval of interest. The forced response to the input is then

$$e(t) = e_0 r(t) + e_1 r'(t) + e_2 r''(t) + \dots \tag{3-99}$$

where $e_0, e_1, \dots$ are called the *error coefficients* of the system. It can be shown that the error coefficients are the coefficients of the Maclaurin series expansion of the error-to-input transfer function $E(s)/R(s)$, i.e.,

$$\frac{E(s)}{R(s)} = e_0 + e_1 s + e_2 s^2 + \dots \tag{3-100}$$

The easiest way to expand $E(s)/R(s)$ for rational functions is to divide the numerator polynomial by the denominator polynomial. The first few error coefficients expressed in terms of the parameters of the open-loop function $C(s)/E(s)$ are listed in Table 3-4 for a unity-feedback system.

Equation (3-99) shows that the response of a system to an arbitrary input can be expressed in terms of the error coefficients, the input, and the derivatives of the input. The only restriction is that the input and its derivatives have no discontinuities in the time interval of interest. In particular, if there is a discontinuity in the input or one of its derivatives, the error-coefficient expansion applies after the transient due to the discontinuity has died out.

In using the expansion, it usually suffices to terminate the series after the first four nonzero error coefficients. If the series is terminated after the kth error coefficient, an approximate bound on the remainder of the series is given by

$$\left|\rho(t)\right| \leq \left|r^{(k+1)}(t)\right|_{\max} \left|e_{k+1}\right| \tag{3-101}$$

where $\rho(t)$ is the remainder, $r^{(k+1)}(t)$ is the $(k+1)$th derivative of the input $r(t)$, and e_{k+1} is the $(k+1)$th error coefficient. The bound [Eq. (3-101)] applies in most practical cases if four or more nonzero terms are used in the expansion. For the restrictions on Eq. (3-101), see reference[40].

TABLE 3-4 ERROR COEFFICIENTS IN TERMS OF OPEN-LOOP FUNCTION $C(s)/E(s)$

Form of Open-Loop Function $C(s)/E(s)$	$\frac{KN(s)}{D(s)}$ $N(s) = n_j s^j + \quad n_2 s^2 + n_1 s + 1$ $D(s) = d_k s^k + \quad + d_1 s + 1$	$\frac{KN(s)}{sD(s)}$	$\frac{KN(s)}{s^2 D(s)}$
e_0	$\frac{1}{1+K}$	0	0
e_1	$\frac{K(d_1 - n_1)}{(1+K)^2}$	$\frac{1}{K}$	0
e_2	$\frac{K}{(1+K)^3}\left\{\begin{array}{l}(d_2 - n_2)(1+K) + \\ (n_1 - d_1)(Kn_1 + d_1)\end{array}\right\}$	$\frac{K(d_1 - n_1) - 1}{K^2}$	$\frac{1}{K}$
e_3	$\frac{K}{(1+K)^4}\left\{\begin{array}{l}(d_3 - n_3)(1+K)^2 + \\ (n_2 - d_2)(1+K) \\ (d_1 + Kn_1) + (n_1 - d_1) \\ (d_2 + Kn_2)(1+K) \\ -(d_1 + Kn_1)^2\end{array}\right\}$	$K^2(d_2 - n_2) + K(2 + Kn_1)$ $(n_1 - d_1) + 1/K^3$	$\frac{d_1 - n_1}{K}$
e_4			$\frac{K(d_2 - n_2) + Kn_1(n_1 - d_1) - 1}{K^2}$

3-8 RESPONSE TO STATIONARY STOCHASTIC INPUTS[48,49,50,51,52]

As discussed in Par. *2-6*, the response of a linear system to a stochastic input cannot be expressed as a specific function of time. The only way to describe system behavior in the presence of stochastic inputs is in terms of the statistics of the input and the response. Theoretically, an infinite number of statistics is required to describe a stochastic process completely. Practically, however, only a few statistics are used.

The probability density functions are direct measures of the chance of occurrence of certain events in the process. The *first probability density function* of a stochastic variable $r(t)$ is denoted by

$P_1(r_1,t_1) \triangleq$ probability that the variable has a value r_1 at time t_1.

The *second probability density function* is denoted by

$P_2(r_1,t_1;r_2,t_2) \triangleq$ probability that the variable has a value r_1 at time t_1 and a value r_2 at time t_2 simultaneously

In practice, only the first two probability density functions are used. For a stationary stochastic process, the first probability density function is independent of the time t_1; the second probability density function is a function only of the time difference $(t_2 - t_1)$.

Two commonly used probability density functions are the *normal distribution* and the *Poisson distribution*. The normal distribution is given by

$$P(r)\,dr = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{r-\bar{r}}{\sigma}\right)^2} dr \qquad \textbf{(3-102)}$$

where $P(r)\,dr$ is the probability of finding r between r and $r + dr$, $\bar{r}$ is the mean value of r (to be defined below), and σ is the standard deviation of r (to be defined below). The Poisson distribution is given by

$$P(N,\Delta t) = \frac{(\nu\Delta t)^N\, e^{-\nu\Delta t}}{N!} \qquad \textbf{(3-103)}$$

where $P(N,\Delta t)$ is the probability of finding N events in a time interval Δt, and ν is the average frequency of occurrence of the events.

In general, the *average* or mean value of a stochastic variable r is given by

$$\bar{r} \triangleq \int_{-\infty}^{+\infty} r\,P(r,t)\,dr \qquad (3\text{-}104)$$

For a stationary stochastic process, the mean value is independent of time and can also be found from

$$\bar{r} \triangleq \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} r(t)\,dt \qquad \textbf{(3-105)}$$

The mean-square value of a stochastic variable or process is given by

$$\overline{r^2} \triangleq \int_{-\infty}^{+\infty} r^2 P(r,t)\,dr \qquad (3\text{-}106)$$

For a stationary stochastic process, the mean-square value is also given by

$$\overline{r^2} \triangleq \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} r^2(t)\,dt \qquad \textbf{(3-107)}$$

The *root-mean-square (rms) value* is the square root of the mean-square value.

The *variance* of a stochastic process if given by

$$V \triangleq \overline{[r-\bar{r}]^2} \qquad \textbf{(3-108)}$$

The *standard deviation* σ is the square root of the variance. It can be expressed in terms of the mean value and the mean-square value as follows:

$$\sigma^2 = \overline{r^2} - \bar{r}^2 \qquad (3\text{-}109)$$

In most applications, rms values and mean values are usually the most common statistics used. To aid in the determination of these quantities, statistics called correlation functions are used. The *autocorrelation function* $\phi_{rr}(\tau)$ of a stationary stochastic process $r(t)$

is defined as the mean value of the product of the function r at time t by the function r at time $t + \tau$, i.e.,

$$\phi_{rr}(\tau) \triangleq \overline{r(t)r(t+\tau)} \tag{3-110}$$

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} r(t)r(t+\tau)\,dt \tag{3-111}$$

The *crosscorrelation* function $\phi_{r\mu}(\tau)$ between two stationary stochastic processes $r(t)$ and $\mu(t)$ is defined as the mean value of the product of the function r at time t by the function μ at time $t + \tau$, i.e.,

$$\phi_{r\mu}(\tau) \triangleq \overline{r(t)\mu(t+\tau)} \tag{3-112}$$

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} r(t)\mu(t+\tau)\,d\tau \tag{3-113}$$

From the definition of the autocorrelation function [Eq. (3-110)], it is evident that the mean-square value of a stochastic process equals the value of the corresponding autocorrelation function with zero argument:

$$\overline{r^2} = \phi_{rr}(0) \tag{3-114}$$

Useful properties of the correlation functions are as follows:

(a) $\phi_{rr}(\tau) = \phi_{rr}(-\tau)$ [even function] (3-115)

(b) $|\phi_{rr}(\tau)| \leq \phi_{rr}(0)$ (3-116)

(c) $\lim_{\tau\to\infty} \phi_{rr}(\tau) = \overline{r}^2$ (3-117)

(d) $\phi_{r\mu}(\tau) = \phi_{\mu r}(-\tau)$ (3-118)

(e) $|\phi_{r\mu}(\tau)| \leq \sqrt{\phi_{rr}(0)\phi_{\mu\mu}(0)}$ (3-119)

(f) $\lim_{\tau\to\infty} \phi_{r\mu}(\tau) = \overline{r}\,\overline{\mu}$ (3-120)

A few examples illustrating the use of autocorrelation functions follow. If $r(t)$ is a rectangular wave with values $+\beta$ or $-\beta$ and with zero crossings located at event points that are Poisson-distributed in time with an average frequency of ν, the autocorrelation function of the process is given by

$$\phi_{rr}(\tau) = \beta^2 e^{-2\nu|\tau|} \tag{3-121}$$

If $r(t)$ is a rectangular wave with amplitude values distributed in any fashion and with zero crossings located at event points Poisson-distributed in time with an average frequency ν, the autocorrelation function of the process is given by

$$\phi_{rr}(\tau) = \sigma^2 e^{-\nu|\tau|} + \overline{r}^2 \tag{3-122}$$

where σ is the standard deviation of the amplitude distribution, and $\overline{r}$ is the mean value of the amplitude distribution.

If $r(t)$ is a train of identical finite pulses whose starting points are Poisson-distributed in time with average frequency ν, the autocorrelation function of the process (known as "shot noise") is given by

$$\phi_{rr}(\tau) = \nu\int_{-\infty}^{+\infty} f(t)f(t+\tau)\,d\tau + \overline{r}^2 \tag{3-123}$$

where $f(t)$ is the time variation or waveform of a single pulse and $\overline{r}$ is given by

$$\overline{r} = \nu\int_{-\infty}^{+\infty} f(t)\,dt \tag{3-124}$$

If $r(t)$ is pure or "white" noise, the autocorrelation function is given by

$$\phi_{rr}(\tau) = \gamma\delta_0(\tau) \tag{3-125}$$

where γ is a constant that depends on how the process is generated.

Thus, if "white" noise is considered as a limiting case of shot noise generated by exponential pulses of amplitude A and time constant T (where the amplitude approaches infinity and the time constant approaches zero with the area s under the pulse held constant), then the constant γ is given by

$$\gamma = \frac{\nu s}{2} \tag{3-126}$$

where ν is the average frequency of occurrence of the pulses.

Because the correlation functions are completely defined as functions of a time variable τ, they are Fourier transformable. By convention, $1/2\pi$ times the Fourier transform of a correlation function is called a power

spectrum or a power-density spectrum. Thus, the *power-density spectrum* $\Phi_{rr}(s)$ of a stochastic process is defined as

$$\Phi_{rr}(s) \triangleq \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\tau\, e^{-s\tau}\, \phi_{rr}(\tau) \qquad (3\text{-}127)$$

The *cross-power-density spectrum* between two stochastic processes $r(t)$ and $\mu(t)$ is defined as

$$\Phi_{r\mu}(s) \triangleq \frac{1}{2k}\int_{-\infty}^{+\infty} d\tau\, e^{-s\tau}\, \phi_{r\mu}(\tau) \qquad (3\text{-}128)$$

Given the power spectra, the corresponding correlation functions can be found by inverse transformation, i.e.,

$$\phi_{rr}(\tau) = \frac{1}{j}\int_{-j\infty}^{+j\infty} \Phi_{rr}(s)\, e^{s\tau}\, ds \qquad (3\text{-}129)$$

$$\phi_{r\mu}(\tau) = \frac{1}{j}\int_{-\infty}^{+j\infty} \Phi_{r\mu}(s) e^{s\tau}\, ds \qquad (3\text{-}130)$$

In terms of the power-density spectrum, the mean-square value of a stochastic process can be found by evaluating the following integral:

$$\overline{r^2} = \int_{-\infty}^{+\infty} \Phi_{rr}(\omega)\, d\omega \qquad (3\text{-}131)$$

Useful properties of the power spectra are

$$\Phi_{rr}(s) = \Phi_{rr}(-s) \text{ (even function)} \qquad (3\text{-}132)$$

$$\Phi_{r\mu}(s) = \Phi_{\mu r}(-s) \qquad (3\text{-}133)$$

Having established some of the statistics of stationary stochastic processes, the response of a linear system to a stochastic input can now be described. If $\phi_{rr}(\tau)$ is the autocorrelation function of the input $r(t)$ of a linear system whose impulse response is $w(t)$, the autocorrelation function of the output $c(t)$ is given by

$$\phi_{cc}(\tau) = \int_{-m}^{+\infty} dt_1\, w(t_1) \int_{-\infty}^{+\infty} dt_2\, w(t_2)\, \phi_{rr}(\tau + t_1 - t_2) \qquad (3\text{-}134)$$

The crosscorrelation function between the input and the output is given by

$$\phi_{rc}(\tau) = \int_{-m}^{+\infty} dt\, w(t)\, \phi_{rr}(\tau - t) \qquad (3\text{-}135)$$

which can be recognized as a convolution integral.

Extending the description of the stochastic response of a linear system to the frequency domain, if $W(s)$ is the transfer function of the system and $\Phi_{rr}(s)$ is the input power-density spectrum, the output power-density spectrum is given by

$$\Phi_{cc}(s) = W(s)\, W(-s)\, \Phi_{rr}(s) \qquad (3\text{-}136)$$

The cross-power-density spectrum between input $r(t)$ and output $c(t)$ is given by

$$\Phi_{rc}(s) = W(s)\, \Phi_{rr}(s) \qquad (3\text{-}137)$$

or

$$\Phi_{cr}(s) = W(-s)\, \Phi_{rr}(s) \qquad (3\text{-}138)$$

If $\mu(t)$ is another signal and $\Phi_{\mu r}(s)$ is the cross-power-density spectrum between $\mu(t)$ and the input $r(t)$, the cross-power-density spectrum between $\mu(t)$ and the output $c(t)$ is given by

$$\Phi_{\mu c}(s) = W(s)\, \Phi_{\mu r}(s) \qquad (3\text{-}139)$$

or

$$\Phi_{c\mu}(s) = W(-s)\, \Phi_{r\mu}(s) \qquad (3\text{-}140)$$

In summary, once the properties of a stochastic process are expressed in terms of correlation functions, the analysis of system behavior is a straightforward problem that can be treated through the use of the definitions and properties of the correlation functions and their transforms, the power spectra. In particular, where rms values are of interest, Eqs. (3-114) and (3-131) are of great use.

3-9 USE OF ANALOG COMPUTERS FOR SIMULATION[52,53,54,55]

In many problems, the use of analog computers greatly facilitates the analysis and design procedures. In both the linear and the nonlinear cases, the analog computer is a tool of wide versatility. Since the detailed properties of analogy computers vary from one manufacturer to the next, this section will cover only some general principles of analog computer use.

The basic elements of any analog computer are integrators, coefficient potentiometers, summing amplifiers, multipliers, and function generators. The symbols for these elements and the mathematical operations they perform are shown in Fig. 3-22. The similarity of these symbols to the symbols of block diagram algebra emphasizes that the block diagram of a system is readily convertible to the computer diagram of the system. Several important restrictions of the computer diagram are

(a) Differentiation is difficult to realize in practice.

(b) The summing amplifiers and integrators almost always introduce a change in algebraic sign.

(c) The useful frequency range of the computer is limited at low frequencies by drift and at high frequencies by phase shift and attenuation.

(d) The amplitude scale of the computer is limited by amplifier saturation.

Example. The block diagram of Fig. 3-7A is shown as a computer diagram in Fig. 3-23.

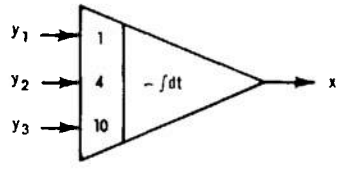
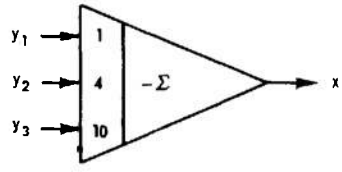
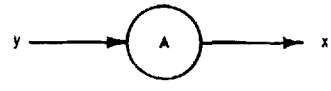
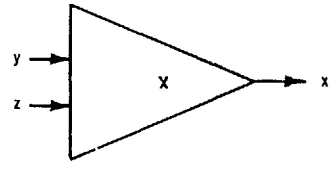
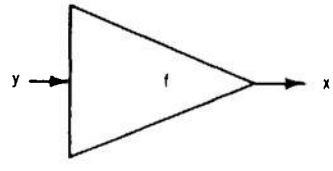
DESCRIPTION	SYMBOL	OPERATION
INTEGRATION		$x = -\int(y_1 + 4y_2 + 10y_3)dt$
ADDITION		$x = -(y_1 + 4y_2 + 10y_3)$
COEFFICIENT		$x = Ay$
MULTIPLICATION		$x = yz$
FUNCTION GENERATION		$x = f(y)$

Fig. 3-22 Elements of analog computers.

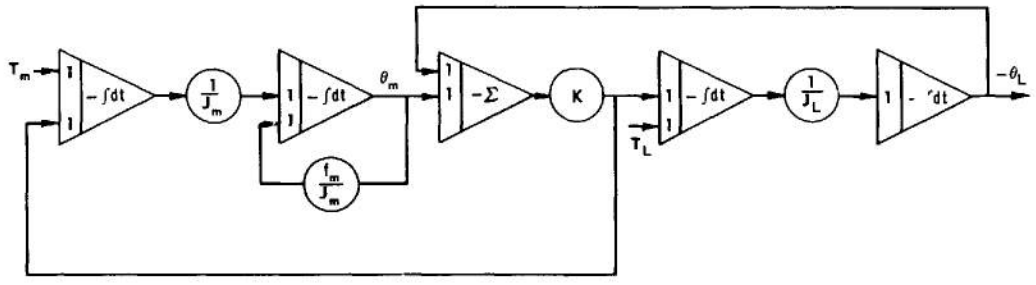


Fig. 3-23 Computer diagram for *system* of Fig. *3-7A.*

BIBLIOGRAPHY

1 P. Franklin, *Methods of Advanced Calculus,* McGraw-Hill Book Company, Inc., New York, N. Y., 1944.

2 H. B. Phillips, *Differential Equations,* John Wiley & Sons, Inc., New York, N.Y., 1945.

3 F. B. Hildebrand, *Advanced Calculus for Engineers,* Prentice-Hall, Inc., New York, N.Y., 1949.

4 E. L. Ince, *Ordinary Differential* Equations, Dover Publications, Inc., New York, N. Y.,' 1944.

5 A. S. Householder, *Principles of Numerical Analysis,* pp. #86-132, McGraw-Hill Book Company, Inc., New York, N.Y., 1953.

6 F. B. Hildebrand, *Introduction to Numerical Analysis,* pp. #451-477, McGraw-Hill Book Company, Inc., New York, N.Y., 1956.

7 H. S. Sharp, "A Comparison of Methods for Evaluating the Complex Roots of Quartic Equations", J. *Math. Phys.,* Vol. 20, No. 3, pp. #243-258, August, 1941.

8 G. S. Brown and D. P. Campbell, *Principles of Servomechanisms,* pp. #89-91, John Wiley & Sons, Inc., New York, N.Y., 1948.

9 S. N. Lin, "A Method of Successive Approximations of Evaluating the Real and Complex Roots of Cubic and Higher-Order Equations", J. *Math. Phys.,* Vol. 20, No. 3, pp. #231-242, August, 1941.

10 S. N. Lin, "A Method for Finding Roots of Algebraic Equations", J. *Math. Phys.,* Vol. 22, pp. #60-77, June, 1943.

11 Y. J. Liu, *Servomechanisms — Charts for Verifying Their Stability and for Finding the Roots of Their Third and Fourth Degree Characteristic Equations,* pp. #55-74, privately printed by Massachusetts Institute of Technology, Elec. Eng. Dept., Cambridge, Mass., 1941.

12 M. F. Gardner and J. L. Barnes, *Transients in Linear Systems,* Vol. 1, John Wiley & Sons, Inc., New York, N. Y., 1942.

13 I. N. Sneddon, *Fourier Transforms,* McGraw-Hill Book Company, Inc., New York, N.Y., 1951.

14 E. A. Guillemin, *The Mathematics of Circuit Analysis,* pp. #517-569, John Wiley & Sons, Inc., New York, N.Y., 1949.

15 G. S. Brown and D. P. Campbell, *Principles of Servomechanisms,* pp. #66-88, 96-98, John Wiley & Sons, Inc., New York, N.Y., 1948.

16 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 1, pp. #66-123, John Wiley & Sons, Inc., New York, N. Y., 1951.

17 W. R. Ahrendt and J. F. Taplin, *Automatic Feedback Control,* pp. #26-52, McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

18 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #4-87, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

19 F. B. Hildebrand, *Advanced Calculus for Engineers,* pp. #52-85, Prentice-Hall, Inc., New York, N. Y., 1949.

20 G. A. Campbell and R. M. Foster, ***Fourier Integrals for Practical Applications,*** D. Van Nostrand Company, Inc., New York, N. Y., **1942**; "Fourier Integrals for Practical Applications", ***Bell System Tech. Pub., Monograph*** **B-584, 1931.**

21 R. V. Churchill, ***Modern Operational Mathematics in Engineering,*** McGraw-Hill Book Company, Inc., New York, N.Y., **1944.**

22 J. **B.** Flannigan and H. S. Kirschbaum, "The Writing of Closed-Loop Control System Transfer Functions by Inspection", ***Trans. AIEE,*** Vol. 71, Part 11, pp. **#90-94, 1952.**

23 T. M. Stout, **"A** Block-Diagram Approach to Network Analysis", ***Trans. AIEE,*** Applications and Industry, Vol. 71, pp. **#255-260, 1952.**

24 T. D. Graybeal, "Block Diagram Network Transformation", ***Elec. Eng.,*** Vol. ***70,*** pp. **#985-990,** November, **1951.**

25 H. Chestnut and R. W. Mayer, ***Servomechanisms and Regulating System Design,*** Vol. **1,** pp. **#194-202,** John Wiley & Sons, Inc., New York, N.Y., **1951.**

26 G. **S.** Brown and D. P. Campbell, ***Principles of Servomechanisms,*** pp. **#140-145,** John Wiley **&** Sons, Inc., New York, N.Y., **1948.**

27 **S.** J. Mason, "Feedback Theory — Some Properties of Signal-Flow Graphs", ***Proc. IRE,*** Vol. **41,** No. **9,** pp. **#1144-1156,** September, **1953.**

28 J. G. Truxal, ***Automatic Feedback Control System Synthesis,*** pp. **#88-160,** McGraw-Hill Book Company, Inc., New York, N.Y., **1955.**

29 W. **R.** Evans, "Control System Synthesis by Root-Locus Method", ***Trans. AIEE,*** Vol. **69,** Part I, pp. **#66-69, 1950.**

30 J. H. Mulligan, Jr., "The Effect **of** Pole and Zero Locations on the Transient Response of Linear Dynamic Systems", ***Proc. IRE,*** pp. **#516-529,** Vol. **37,** May, **1949.**

31 H. Chestnut and R. **W.** Mayer, "Comparison of Steady-State and Transient Performance of Servomechanisms", ***Trans. AIEE,*** Vol. **68,** Part I, pp. **#765-777, 1949.**

32 H. Harris, Jr., M. J. Kirby, and E. F. von Arx, "Servomechanism Transient Performance from Decibel-Log Frequency Plots", ***Trans. AIEE,*** Vol. ***70,*** Part 11, pp. **#1452-1459, 1951.**

33 T. J. Higgins and J. G. Levinthal, "Stability Limits for Third-Order Servomechanisms", ***Trans. AIEE,*** Vol. **71,** Part 11, pp. **#459-467, 1952.**

34 G. Biernson, "Quick Methods for Evaluating the Closed-Loop Poles of Feedback Control Systems", ***Trans. AIEE,*** Vol. **72,** Part 11, pp. **#53-70, 1953.**

35 C. H. Dawson, "Approximation of Transient Response from Frequency Response Data", ***Trans. AIEE,*** Vol. **72,** Part 11, pp. **#289-291, 1953.**

36 A. H. Zemanian, "Bounds Existing on the Time and Frequency Responses of Various Types of Networks", ***Proc. IRE,*** Vol. **42,** pp. **#835-839,** May, **1954.**

37 A. H. Zemanian, "Further Effects of the Pole and Zero Locations **on** the Step Response of Fixed, Linear Systems", ***Trans. AIEE,*** Vol. **74,** Part 11, pp. **#52-55, 1955.**

38 D. V. Stallard, "A Series Method of Calculating Control System Transient Response from the Frequency Response", ***Trans. AIEE,*** Vol. **74,** Part 11,pp. **#61-64, 1955.**

39 **H.** Thal-Larsen, "Frequency Response from Experimental Nonoscillatory Transient-Response Data", ***Trans. AIEE,*** Vol. **74,** Part **11,** pp. **#109-114, 1955.**

40 G. A. Biernson, "A Simple Method for Calculating the Time Response of a System to an Arbitrary Input", *Trans. AIEE,* Vol. 74, Part 11, pp. #227-245, 1955.'

41 G. S. Brown and D. P. Cambell, *Principles* of *Servomechanisms,* pp. #332-365, John Wiley & Sons, Inc., New York, N. Y., 1948.

42 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 1, pp. #398-439, John Wiley & Sons, Inc., New York, N. Y., 1951.

43 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 2, pp. #25-35, John Wiley & Sons, Inc., New York, N.Y., 1955.

44 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #61-76, 375-390, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

45 E. A. Guillemin, "Computational Techniques Which Simplify the Correlation Between Steady-State and Transient Responses of Filters and Other Networks", *Proc. Nat. Electronics Conf. 1953,* Vol. 9, 1954.

46 J. L. Bower, "A Note on the Error Coefficients of a Servomechanism", *J. Appl. Phys.,* Vol. 21, p. #723L, July, 1950.

47 E. Arthurs and L. H. Martin, "Closed Expansion of the Convolution Integral (A Generalization of Servomechanism Error Coefficients)", *J. Appl. Phys.,* Vol. 26, pp. #58-60, January, 1955.

48 N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications,* John Wiley & Sons, Inc., New York, N.Y., 1949.

49 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms,* MIT Radiation Laboratory Series, Vol. 25, pp. #262-368, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

50 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 2, pp. #65-88, John Wiley & Sons, Inc., New York, N. Y., 1955.

51 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #410-499, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

52 J. H. Laning, Jr. and R. H. Battin, *Random Processes in Automatic Control,* McGraw-Hill Book Company, Inc., New York, N.Y., 1956.

53 Edited by I. A. Greenwood, Jr., J. V. Holdam, Jr., and I. Macrae, Jr., *Electronic Instruments,* MIT Radiation Laboratory Series, Vol. 21, Part 1, McGraw-Hill Book Company, Inc., New York, N.Y., 1948.

54 A. C. Hall, "A Generalized Analog Computer for Flight Simulation", *Trans. AIEE,* Vol. 69, Part 11, pp. #308-320, 1950.

55 R. A. Bruns, "Analog Computers for Feedback Control Systems", *Trans. AIEE,* Vol. 71, Part 11, pp. #250-254, 1952.

CHAPTER 4

STABILITY OF FEEDBACK CONTROL SYSTEMS*

4-1 INTRODUCTION

The determination of system stability is the first step in the design of any linear control system. To carry out this first step, a test for system stability is required. The particular stability test used will depend on the meaning attached to the term stable operation. Generally, a system is said to be stable if it remains at rest when all inputs are zero and if (for any disturbance) no signal grows without bound or exhibits sustained oscillation when the inputs are returned to zero. In the case of linear systems, the only situation in which unstable behavior can occur is the one in which the roots of the characteristic equation of the closed-loop system lie in the right-half s plane and therefore have positive real parts. The response modes corresponding to right-half-plane roots of the characteristic equation have amplitudes that increase without limit as time increases. Consequently, any stability criterion for a linear system is essentially a method of determining whether or not the characteristic equation has right-half-plane roots.

In Fig. **4-1** the general single-loop system is shown. The output response transform for this system is

$$C(s) = \frac{G_1(s)\,G_2(s)\,R(s) - G_2(s)\,U(s)}{1 + G_1(s)\,G_2(s)\,H(s)} \qquad (4\text{-}1)$$

The characteristic equation of the system is

$$1 + G_1(s)\,G_2(s)\,H(s) = 0 \qquad (4\text{-}2)$$

Thus, if any of the roots of Eq. **(4-2)** lie in the right-half s plane, the system of Fig. **4-1** is unstable.

The presence of right-half-plane roots of the denominator of the response transform $C(s)$ [i.e., right-half-plane poles of $C(s)$] could be determined by direct factorization of Eq. (4-2), after it has been cleared of fractions. Since a system is unstable if ***one or more*** right-half-plane poles of $C(s)$ exist, it is usually sufficient to determine whether these poles exist; however, it is not necessary to determine their exact location. Hence, the standard stability criteria that are discussed in this chapter (with the exception of the root-locus method) merely determine the number of unstable poles without regard to their location in the right-half s plane.

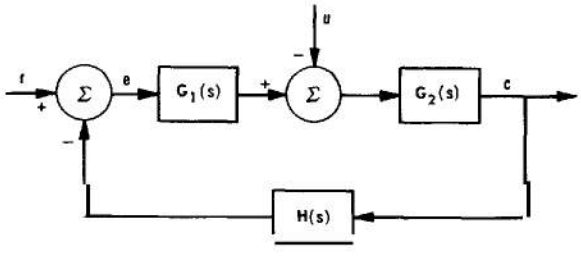


Fig. 4-1 Single-loop system — block diagram.

By L. A. Gould

Three useful stability criteria —the Routh criterion, the Nyquist criterion, and root-locus method —are described in succeeding paragraphs.

The Routh criterion is the simplest to apply and can be used when the characteristic equation is known, at least in literal form. This criterion can be employed to determine the presence of any roots of an algebraic equation that lie in the right-half plane. If the equation is the characteristic equation of a closed-loop system, the presence of right-half-plane roots means that the system is unstable.

The Nyquist criterion is the most widely used stability criterion because the only information it requires for its application is a plot of the open-loop frequency response $G_1(j\omega)\ G_2(j\omega)\ H(j\omega)$. This frequency response can be determined either from an analytical representation of component behavior or from direct measurement of the response of the components to sinusoidal inputs. In addition to its use in determining the presence of system stability, the Nyquist criterion is extended in many design procedures to give an indication of the ***degree of stability*** possessed by the stable system (see Chs. 5 and 6). By an examination of the behavior of the $G_1(j\omega)\ G_2(j\omega)\ H(j\omega)$ locus in the vicinity of the $-1+j0$ point, the Nyquist criterion provides the servo engineer with a relatively straightforward and extremely powerful tool for analysis and design.

The root-locus method is a graphical technique for revealing the position of the poles of the response transform $C(s)$ in the s plane as a gain factor of the open-loop function $G_1(s)\ G_2(s)\ H(s)$ is varied. The primary advantage of this method for stability determination is that the closed-loop pole locations are kept in evidence at all times. Thus, it is easy to see when the poles move into the right-half plane as the gain factor is varied. There are two primary disadvantages connected with the root-locus method. First, the location of the poles *and* zeros of the open-loop function must be specified. This often requires some sort of analytical approximation to the experimental test data. Second, the plotting of the paths of the closed-loop poles involves a trial-and-error procedure that can be quite tedious. In spite of these disadvantages, however, the root-locus method is quite useful in that it immediately places in evidence the closed-loop pole-zero configuration for any particular design (stable, of course). Thus, the characteristics of the time response of the system are easily ascertained and the verification of performance specifications in the time domain is a straightforward matter.

4-2 ROUTH CRITERION[1,2,6]

By applying the Routh stability criterion, one can determine whether any roots of an algebraic equation lie in the right-half s plane. If the coefficients of the equation are known only in literal form, the Routh criterion yields only a set of inequality conditions for stability. However, if the coefficients of the equation are known numerically, the criterion permits one to determine whether stability actually exists.

To show the general procedure used in applying the Routh criterion, consider the following general algebraic equation :

$$a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0 = 0 \qquad (4\text{-}3)$$

Next, the coefficients are arrayed in two rows, alternate coefficients being placed in alternate rows

$$\begin{array}{lccccc} (1) & a_n & a_{n-2} & a_{n-4} & \ldots & 0 \\ (2) & a_{n-1} & a_{n-3} & a_{n-5} & \ldots & 0 \end{array} \qquad (4\text{-}4)$$

Then, the array is extended by taking appropriate cross-products to determine the elements in the third row

$$(3)\quad \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} \quad \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}} \quad \ldots \qquad (4\text{-}5)$$

The elements of the fourth row are formed by taking cross-products of the elements of the second and third rows, in exactly the same manner that the third-row elements were formed. This process is continued until all the elements of a row are zero. On completion of the array, the Routh criterion can be employed to determine the presence of right-half-plane roots.

The Routh criterion states:

> The number of roots of the original equation that lie in the right-half s plane equals the number of sign changes in the elements that form the first column of the final array.

An examination of the procedure used above shows that all the elements of any row after the second may be divided by a positive number without changing the result.

Examples.

(a) Consider the following algebraic equation with lateral coefficients:

$$s^4 + Ks^3 + 2s^2 + 4s + M = 0 \qquad (4\text{-}6)$$

where $K > 0$ and $M > 0$.

The complete algebraic array is as follows:

(1)	1	2	M	0		
(2)	K	4	0	0		
(3)	$(K\text{-}2)$	$KM/2$	0	0	(row multiplied by $K/2$)	(4-7)
(4)	$[(K\text{-}2) - K^2M/8]$	0	0	0	(row multiplied by 1/4)	
(5)	$(K\text{-}2\text{-}K^2M/8)$	0	0	0	(row multiplied by $2/KM$)	
(6)	0	0	0	0		

The inequalities that determine stability are

$$K > 0, \qquad (4\text{-}8)$$

$$K\text{-}2 > 0, \qquad (4\text{-}9)$$

and

$$[(K\text{-}2) - K^2M/8] > 0 \qquad (4\text{-}10)$$

(b) As an example of an algebraic equation with numerical coefficients, consider the equation

$$s^5 + 2s^4 + 2s^3 + 46s^2 + 89s + 260 = 0 \qquad (4\text{-}11)$$

The complete numerical array is as follows:

(1)	1	2	89	0		
(2)	2	46	260	0		
(3)	−1	−1.95	0	0	(row divided by 21)	(4-12)
(4)	1	6.17	0	0	(row divided by 42.1)	
(5)	1	0	0	0	(row divided by 4.22)	
(6)	1	0	0	0	(row divided by 6.17)	
(7)	0	0	0	0		

Inspection of the signs of the elements of the first column shows that one sign change occurs in going from the second to the third row and another in going from the third to the fourth row. Hence, two roots of Eq. (4-11) lie in the right-half s plane. The factors of Eq. (4-11) are

$$(s+4)(s-2+j3)(s-2-j3)$$
$$(s+1+j2)(s+1-j2) \qquad (4\text{-}13)$$

The right-half s-plane roots of Eq. (4-13) are

$$s_1, s_2 = +2 \pm j3 \qquad (4\text{-}14)$$

4-3 NYQUIST CRITERION[3,6]

The Nyquist criterion is a graphical procedure by which one can determine whether any of the roots of the equation

$$1 + G(s) = 0 \qquad (4\text{-}15)$$

lie in the right-half s plane. Only the following information is required in this procedure: (1) the magnitude and phase angle of $G(j\omega)$; (2) the behavior of G(s) at the poles of G(s) that lie on the imaginary axis or at the origin of the s plane; and (3) the number of poles of G(s) in the right-half s plane. (NOTE: For nonunity feedback loops, one tests for the zeros of the function $1 + G(s)H(s)$ where $G(s)$ is the forward transfer function and $H(s)$ is the feedback transfer function.)

The Nyquist criterion can be expressed mathematically as

$$Z = N + P \qquad (4\text{-}16)$$

where

Z = number of zeros of $1 + G(s)$ that lie in the right-half s plane

P = number of poles of G(s) that lie in the right-half s plane

N = number of clockwise encirclements of the point $-1 + j0$ by the locus of G(s) as s describes the path shown in Fig. 4-2

For stability, Z must be zero; that is, $P = -N$. If $P \neq -N$, the system is unstable.

If there are any poles of G(s) on the imaginary axis, the $G(j\omega)$ locus will become infinite at these points. To determine the behavior of the $G(j\omega)$ locus at these poles, so as to be able to count encirclements, a small

semicircular detour is made into the right-half s plane at each pole of $G(s)$ on the imaginary axis. Thus, the $G(j\omega)$ locus will describe a large semicircle instead of becoming infinite. If the pole on the imaginary axis is of multiple order, the $G(j\omega)$ locus will describe one semicircle for each order of the multiple pole.

A convenient rule for determining the direction of turn of the $G(j\omega)$ locus at the imaginary-axis poles of G(s) is

Turn to the *right* by 180° for each order of the pole as the frequency increases.

If $G(s)$ has no poles in the right-half s plane or on the imaginary axis (except at the origin), the Nyquist stability criterion simplifies to†

$$|G(j\omega)| < 1 \qquad \text{when } \angle\, G(j\omega) = -180° \tag{4-17}$$

Examples.

(a) Consider the function

$$G(s) = \frac{K}{s(Ts+1)^2} \tag{4-18}$$

For what range of K will $1 + G(s)$ have stable roots?

†Symbol $\angle$ denotes "angle"

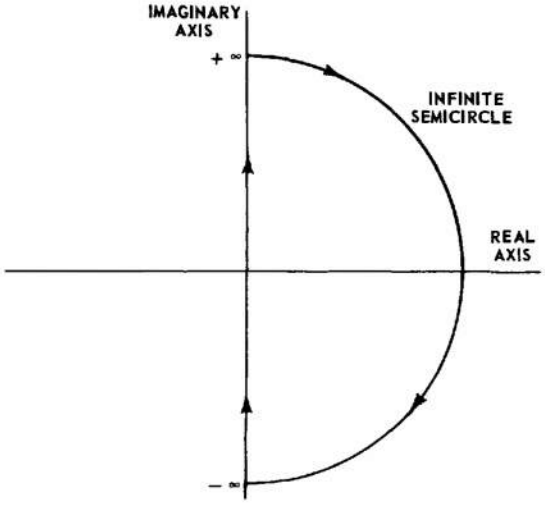


Fig. 4-2 locus of s for the Nyquist criterion.

To simplify the calculation, change variables, letting $\lambda = Ts$. Then,

$$G(\lambda) = \frac{KT}{\lambda\,(\lambda+1)^2} \tag{4-19}$$

Plot $(1/KT)\,G(\lambda)$ on the complex plane for $\lambda = ju$. Such a plot is sketched in Fig. 4-3. Since encirclements depend *only* upon the topology of the plot, the locus can be distorted to facilitate the counting of encirclements (Fig. 4-4).

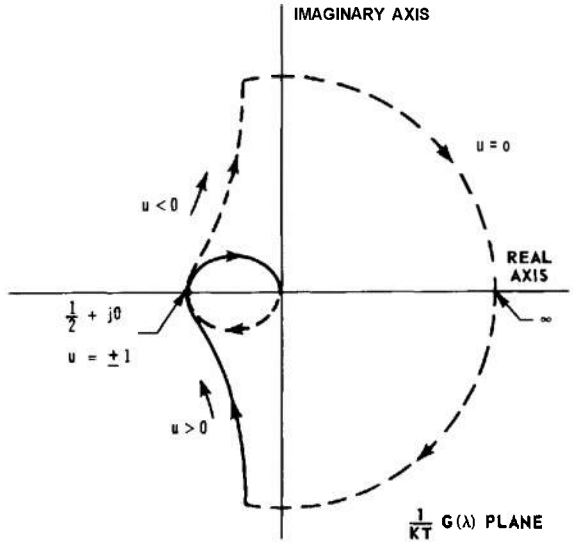


Fig. 4-3 Locus of $\frac{1}{\lambda(\lambda+1)^2}$ for $\lambda = j\mu$.

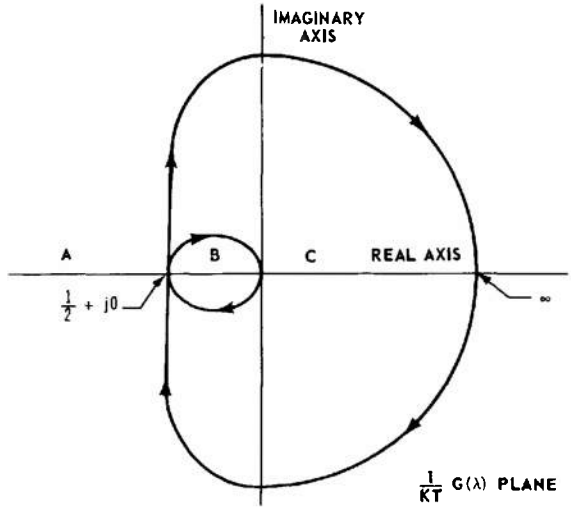


Fig. 4-4 Distortion of locus of $\frac{1}{h(h+1)^2}$.

It is convenient to arrange the pertinent information in tabular form as follows:

Location of $-1+j0$	P	N	Z	Nature of Stability
A	0	0	0	stable ($Z=0$)
B	0	2	2	unstable ($Z \neq 0$)
C	0	1	1	unstable ($Z \neq 0$)

The stable range of KT is determined from the table above and **Eq. (4-17)** as

$$0 < KT < 2.0$$

This expression is found by determining the range of KT which, when multiplying the $(1/KT)G(\lambda)$ locus, will keep the point $-1+j0$ in region A.

(b) Consider the function

$$G(s) = K\frac{(1+s)}{s(1-s)} \qquad \text{(4-20.)}$$

For what range of K will $1+G(s)$ have stable roots?

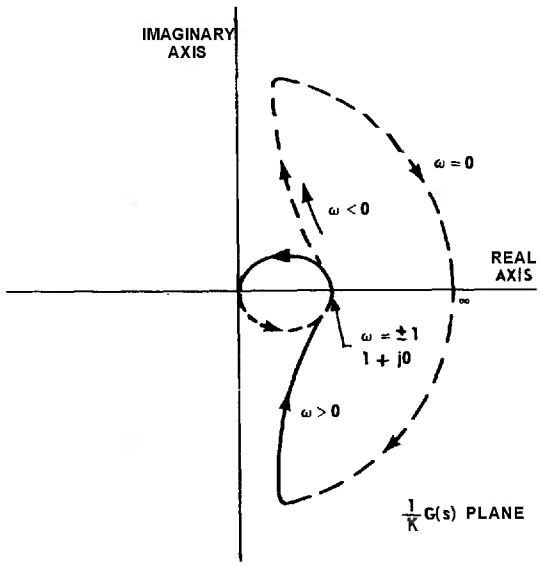


Fig. 4-5 Locus of $\frac{1}{s}\left(\frac{1+s}{1-s}\right)$.

Location of $-1+j0$	P	N	Z	Nature of Stability
A	1	0	1	unstable
B	1	−1	0	stable
C	1	1	2	unstable

Stable roots exist for K in the range

$$-\infty < K < -1$$

These limits are found by determining the range of K which, when multiplying the $(1/K)\,G(s)$ locus, will keep the $-1+j0$ point in region B.

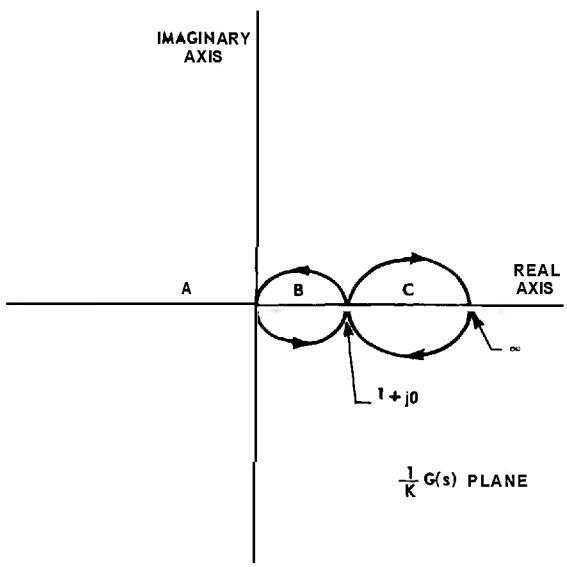


Fig. 4-6 **Distortion of locus of** $\frac{1}{s}\left(\frac{1+s}{1-s}\right)$.

4-4 ROOT-LOCUS METHOD[4,5]

In the root-locus method, a plot is made of the locus of the roots of $1 + G(s) = 0$ as a function of a gain factor of $G(s)$. $G(s)$ must be known numerically in completely factored form. For nonunity feedback loops, one plots the roots of the equation $1 + G(s)H(s) = 0$ where $G(s)$ is the forward transfer function and $H(s)$ is the feedback transfer function.

If $G(s)$ is written in the form

$$G(s) = K_o \frac{\prod_{k=1}^{z} (s + s_k)}{\prod_{j=1}^{p} (s + s_j)} \qquad (4\text{-}21)$$

where K_o is varied from 0 to $+\infty$, then the necessary condition for a point in the s plane to lie on the locus of the roots of $1 + G(s) = 0$, as K_o varies, is

$$\sum A, - \sum A, = -180° \qquad (4\text{-}22)$$

where

$\sum A,$ = sum of the angles of the phasors from the ***zeros*** of $G(s)$ to the point in question

$\sum A,$ = sum of the angles of the phasors from the ***poles*** of $G(s)$ to the point in question

The value of the constant K_o that is associated with each root-locus point is found from the relation

$$K_o = \frac{\prod |V_p|}{\prod |V_z|} \qquad (4\text{-}23)$$

where

$\prod |V_p|$ = product of the magnitudes of the phasors from the ***poles*** of $G(s)$ to the root-locus point

$\prod |V_z|$ = product of the magnitudes of the phasors from the ***zeros*** of $G(s)$ to the root-locus point

The root-locus method can be used to reveal the position of the roots of $1 + G(s) = 0$ directly and to determine whether any of the roots can move into the right-half s plane as the constant K_o is varied. This method of stability determination is primarily a graphical one, particularly when determining the points in the s plane that satisfy the angle condition [Eq. (4-22)]. Although the angle condition determines the entire locus, it is still necessary to find the actual points by a trial-and-error procedure. That is, a point is guessed and the angle condition is checked; if the angle condition is not satisfied, another point is tried, etc.

To facilitate the plotting of the root locus, several theorems based on the angle condition [Eq. (4-22)] and the magnitude condition [Eq. (4-23)] have been established. These are:

(a) The number of branches for a given locus equals the number of roots of $1 + G(s) = 0$.

(b) The locus starts ($K_o = 0$) at poles and ends ($K_o = \infty$) at zeros.

(c) The real-axis position of the locus always has an odd number of poles and zeros to the right of the s point for $K_o > 0$.

(d) The breakaway from the real axis into the complex plane between two adjacent poles occurs at the point of maximum K_o.

(e) For two adjacent zeros, the locus enters the real axis from the complex plane at the point of minimum K_o.

(f) Near complex ***poles***, the direction of the locus is given by

$$[180° - \sum A_z + \sum A_p]$$

where $\sum A,$ is the sum of the angles of the phasors from all the other ***zeros*** to the complex pole in question, and $\sum A,$ is the sum of the angles of the phasors from all the other ***poles*** to the complex pole in question. Near complex ***zeros***, the direction of the locus is given by

$$[-180° + \sum A_p - \sum A_z].$$

(g) The asymptotes of the locus for large values of s are given by a set of straight lines that intersect the real axis at angles

$$A = \frac{180° \pm 360°n}{p - z} \qquad (n=0, 1, 2, \ldots) \tag{4-24}$$

and whose intersection with the real axis is given by the centroid of the pole-zero configuration

$$x_o = \frac{\sum_{j=1}^{p} s_j - \sum_{k=1}^{z} s_k}{p - z} \tag{4-25}$$

where

p = number of poles $\qquad s_j = j^{th}$ pole

z = number of zeros $\qquad s_k = k^{th}$ zero

(h) The locus is symmetrical with respect to the real axis.

These theorems may be verified for the various loci in Fig. 4-7, which presents examples of a wide variety of root loci for systems up to the fourth order. In this figure, $T(p, x)$ indicates a system with p poles and z zeros.

As an example of a typical root-locus plot for a unity-feedback loop, consider the function

$$G(s) = \frac{K}{s(s/\omega_1 + 1)(s/\omega_2 + 1)} = \frac{K\omega_1\omega_2}{s(s+\omega_1)(s+\omega_2)}$$

where $\omega_1 = 10$ rad/sec, $\omega_2 = 30$ rad/sec, and the conditions are

$$/s + /s + 10 + /s + 30 = 180° \quad \text{(angle condition)}$$

$$|s||s + 10||s + 30| = 300K \quad \text{(magnitude condition)}$$

The location of the poles of $G(s)$ on the s plane is as shown in Fig. 4-8.

The root locus coincides with the real axis lying between the pole at the origin and the pole at -10, as well as with the part of the real axis lying to the left of the pole at -30. The locus breaks away from the real axis at some point between the pole at the origin and the pole at -10. To locate this breakaway point, either the technique described in theorem (d) or, in this simple case, an analytical technique can be applied.

Let -6 be the location of the breakaway point. With $s = -6$, the magnitude condition becomes

$$f(\delta) = 6(-6 + 10)(-6 + 30) = 300K$$

The negative of the value of 6 which maximizes the left side of the equation above is the coordinate of the breakaway point. Thus

$$f(\delta) = \delta^3 - 40\delta^2 + 300\delta$$

$$f'(\delta) = 3\delta^2 - 80\delta + 300 = 0$$

$$\delta = 20.8, 6.3$$

Only the smaller value of 6 satisfies the angle condition on the real axis. Therefore, the breakaway point is at -6.3. The gain factor K at this point is obtained by substituting this value of 6 into the magnitude condition equation, i.e.,

$$300K = 6.3(-6.3 + 10)(-6.3 + 30)$$

$$K = 1.84$$

The point on the branch of the locus to the left of the pole at -30, for this same value of gain K, is at -30.8.

The asymptotes of the locus, for large values of s, intersect the real axis at angles given by the relation

$$A = \frac{180° \pm 360° n}{p - z}$$

Since $p = 3$ and $x = 0$, the angles are $60°$ and $-60°$ (or $+120°$). The real-axis intercept of the asymptotes is given by

$$x_o = \frac{\sum s_j - \sum s_k}{p - z} = \frac{[0 + 10 + 30] - [0]}{3 - 0} = 13.3$$

A sketch of the data obtained so far is given in Fig. 4-9.

At this point, it would seem necessary to apply the exploratory s-point method to determine the rest of the locus. However, even before this is done, we can determine which of the asymptotes the locus approaches. In this particular case, we already know (from a Nyquist plot) that the function $1 + G(s)$ will have right-half-plane roots when the sensitivity is above a certain value. Therefore, the two branches of the locus in the complex portion of the plane must head toward the right-half plane as the gain factor K increases. Thus, the locus will cross the imaginary axis and *go* into the right-half plane. The points of imaginary-axis crossing are easy to find in this particular case because of the small number of poles involved in the configuration.

Consider the geometric properties of Fig. **4-10.** At the crossing point, the angle condition requires that

$$\tan^{-1}\frac{\omega_c}{\omega_1} + \tan^{-1}\frac{\omega_c}{\omega_2} + 90° = 180°$$

Therefore

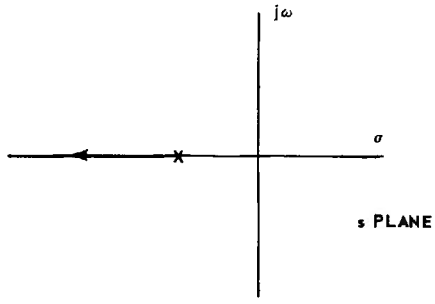
$$\frac{\left(\frac{\omega_c}{\omega_1}\right) + \left(\frac{\omega_c}{\omega_2}\right)}{1 - \left(\frac{\omega_c^2}{\omega_1\,\omega_2}\right)} \to \infty$$

or

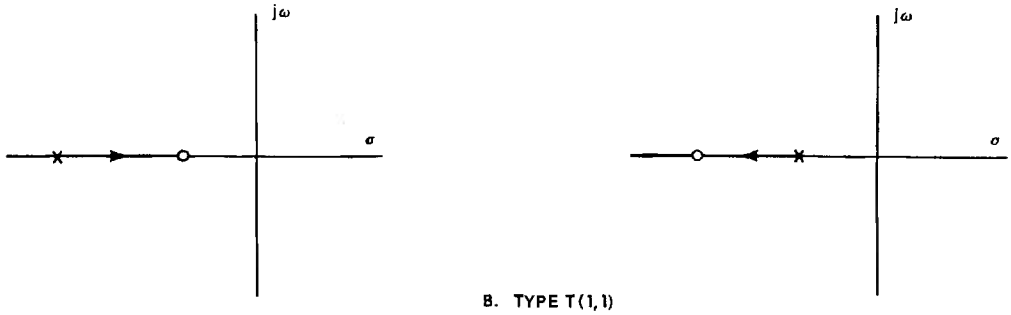
$$\omega_c = \sqrt{\omega_1\,\omega_2} = \sqrt{300} = 17.3$$

At this point, the gain K is determined **by** the magnitude condition. Using this condition, it is found that $K = 40$. The corresponding point on the branch to the left of the pole at -30 is at -40.

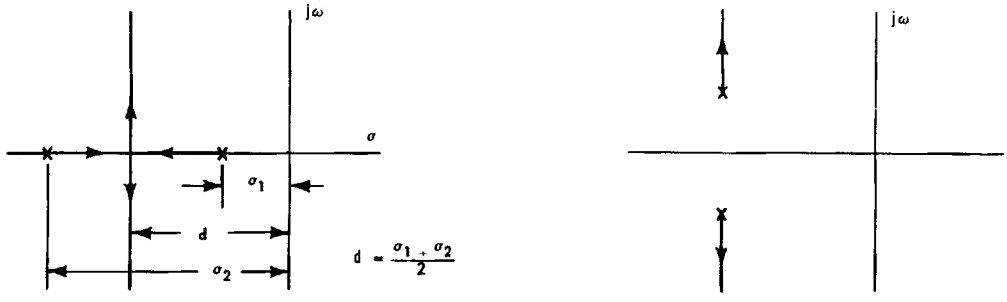
The remainder of the locus can be sketched in, or a more accurate determination of the locus points can be made by the exploratory s-point method. Figure **4-11** is a sketch **of** the entire locus, with key points indicated.



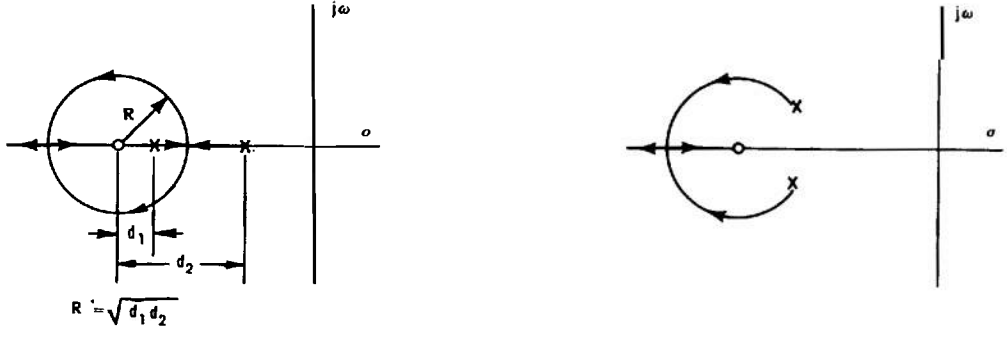
A. TYPE T(1,0)



B. TYPE T(1,1)



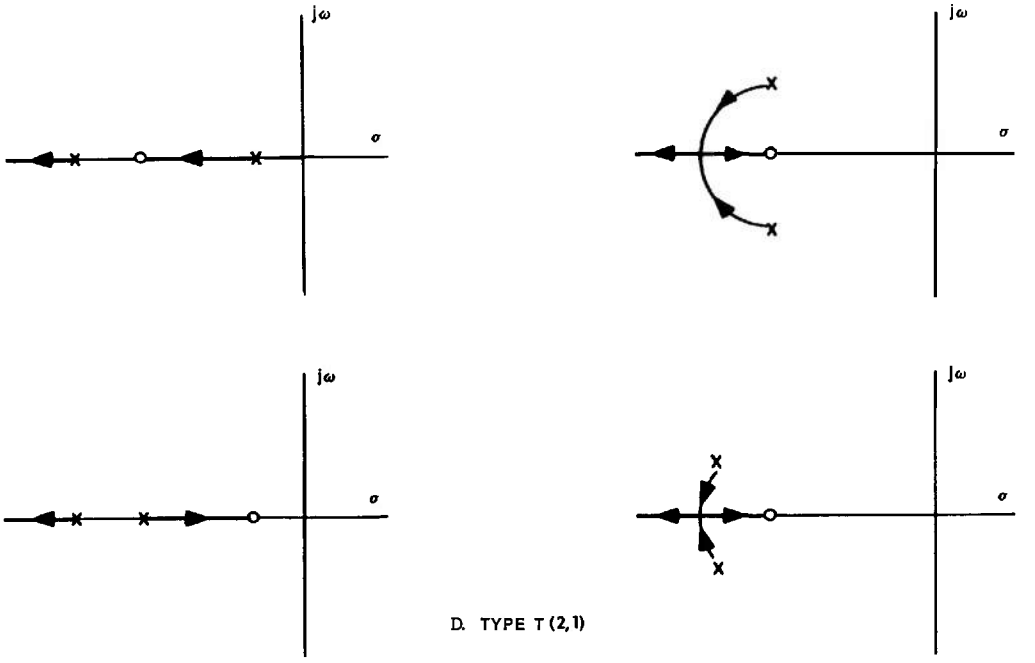
C. TYPE T(2,0)



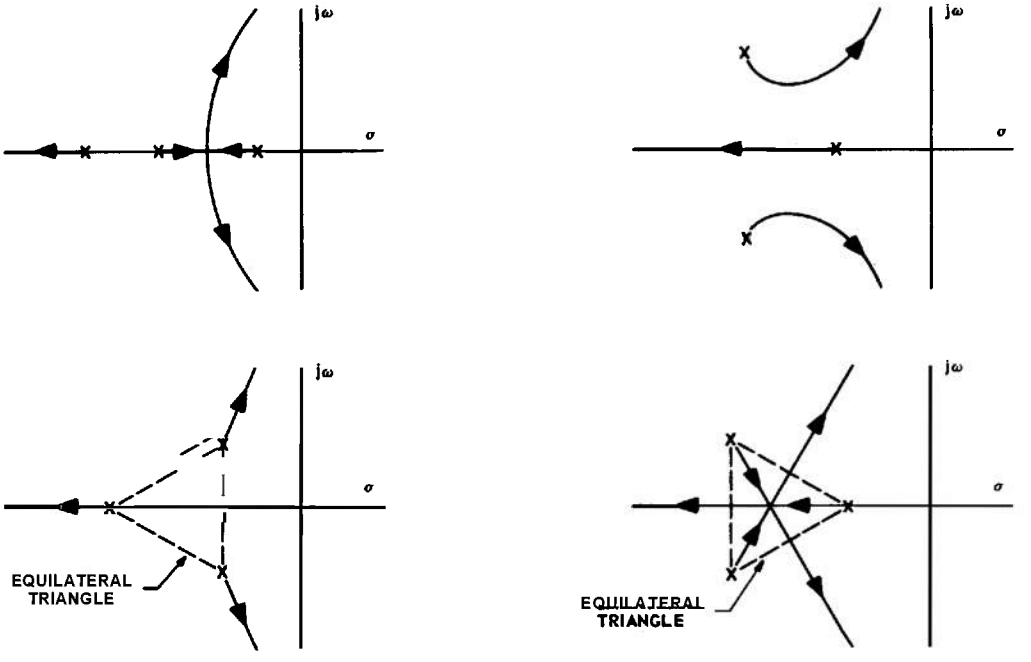
D. TYPE T(2,1)

Fig. 4-7 Root-loci plots. (Sheet 1 of 5)

Adapted from "The Study of Transients in Linear Feedback Systems by Conformal Mapping and the Generalized Root Locus Method", by V. C. M. Yeh, ScD Thesis M.E., 1962, Massachusetts Institute of Technology.

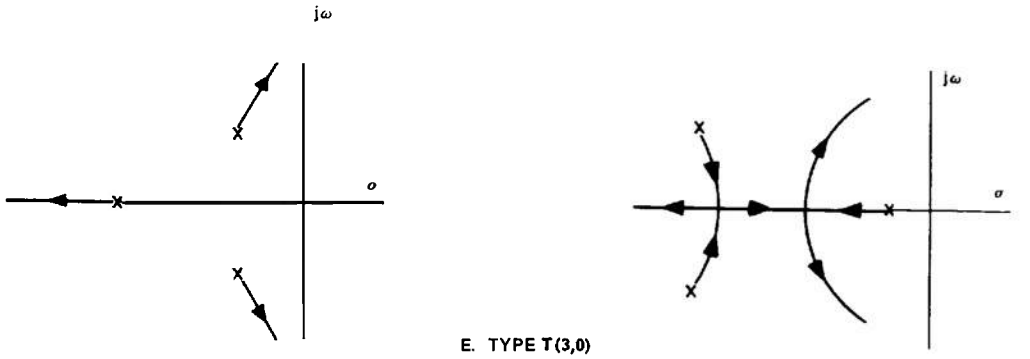


D. TYPE T (2,1)

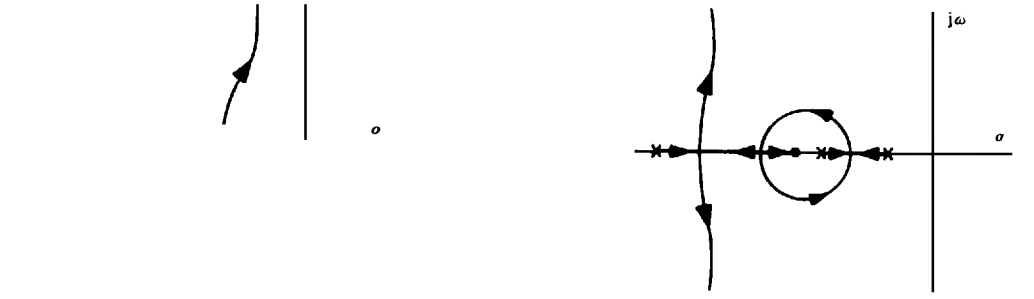


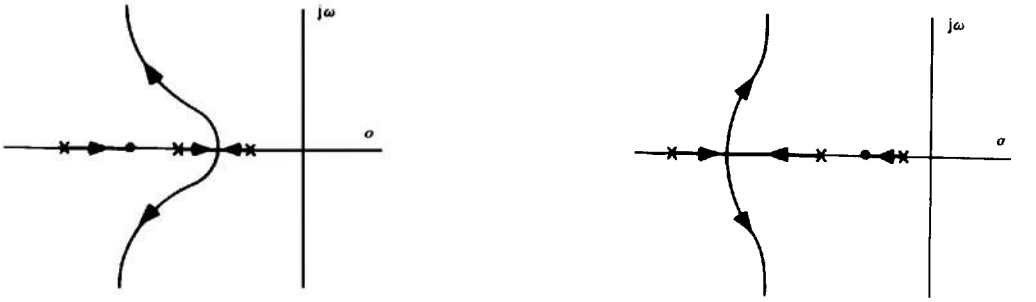
E. TYPE T (3,0)

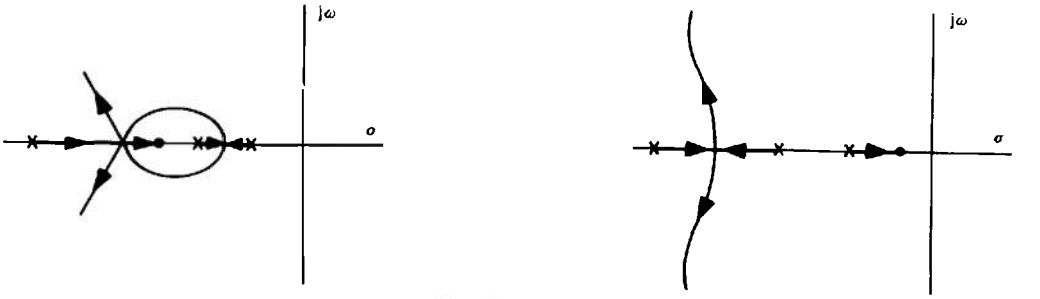
Fig. 4-7 Root-loci plots. (Sheet 2 of 5)



E. TYPE T(3,0)

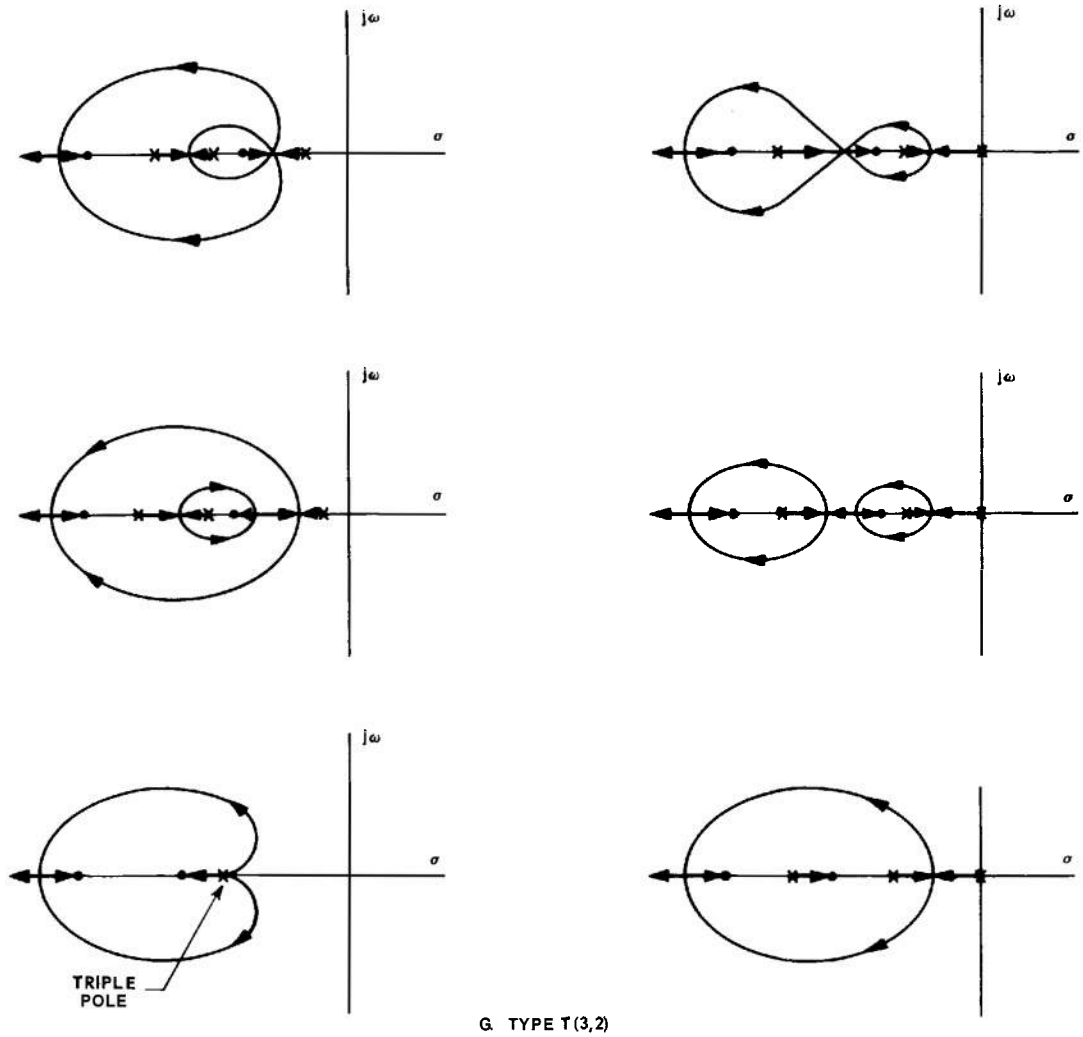




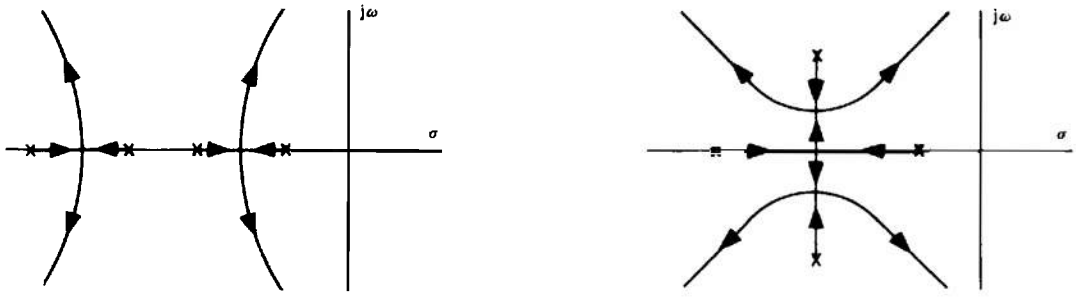


F. TYPE T(3,1)

Fig. 4-7. Root-loci plots. (Sheet 3 of 5)

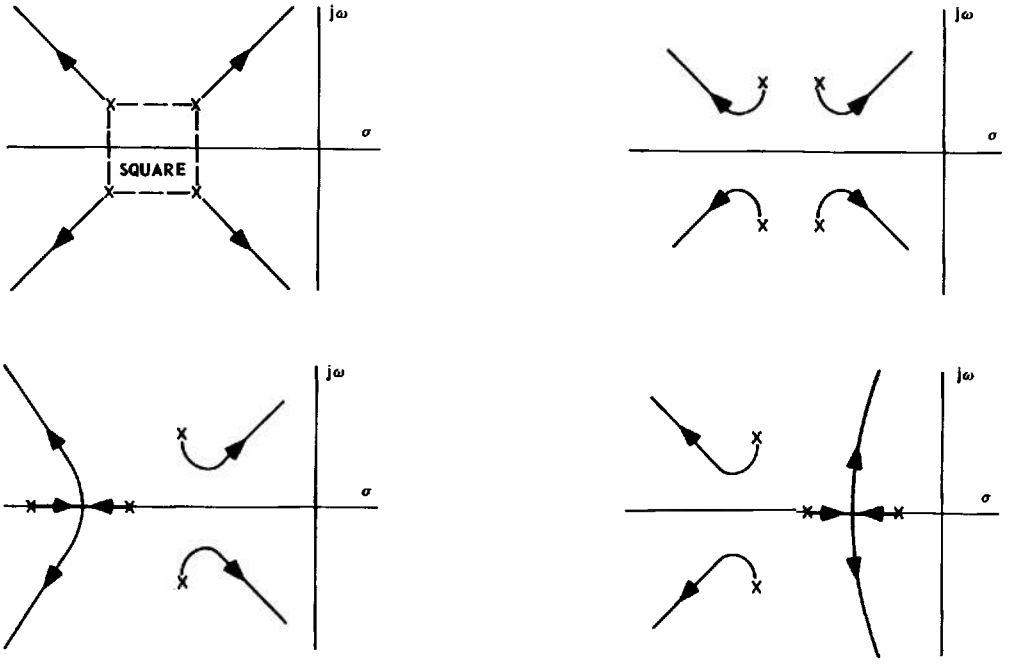


G. TYPE T(3,2)



H. TYPE T(4,0)

Fig. 4-7 Root-loci plots. (Sheet 4 of 5)



H. TYPE T (4,0)

Fig. 4-7 Root-loci plots. (Sheet *5 of 5)*

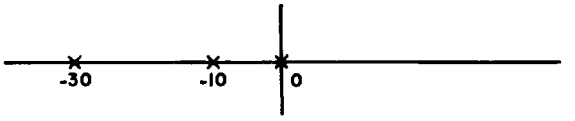


Fig. 4-8 Poles for $G(s) = \dfrac{300K}{s(s+10)(s+30)}$.

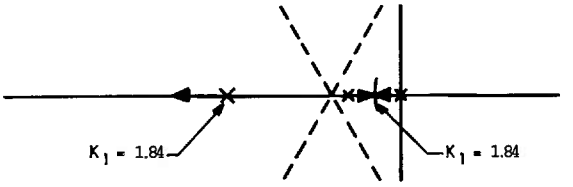


Fig. 4-9 Asymptotes and real axis behavior for

$$G(s) = \frac{300K}{s(s+10)(s+30)}.$$

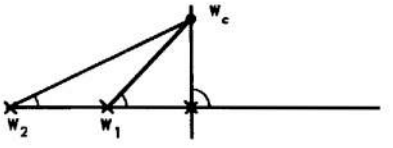


Fig. 4-10 Construction for determining imaginary axis crossing $G(5) = \dfrac{300K}{s(s+10)(s+30)}$ *when* $0 < K < \infty$.

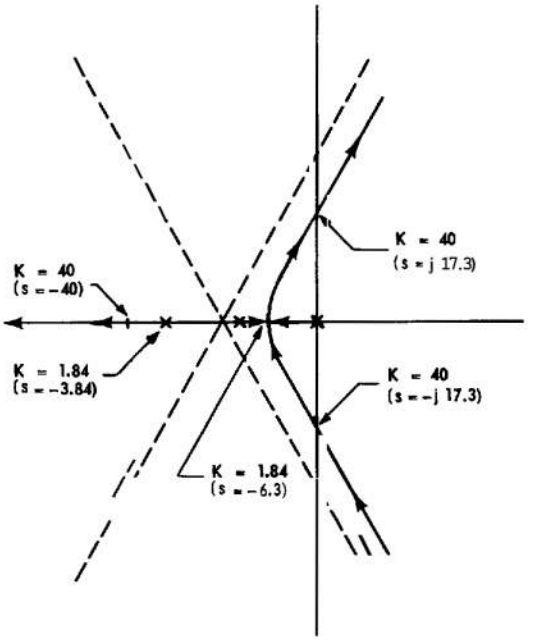


Fig. 4-11 Complete root locus for $G(s) = \dfrac{300K}{s(s+10)(s+30)}$ *when* $0 < K < \infty$.

BIBLIOGRAPHY

1 **M.** F. Gardner and J. L. Barnes, *Transients in Linear Systems,* Vol. I, pp. #197-201, John Wiley & Sons, Inc., New York, N.Y., **1942.**

2 W. R. Ahrendt and J. F. Taplin, *Automatic Feedback Control,* pp. #78-80, McGraw-Hill Book Company, Inc., New York, N.Y., **1951.**

3 H. W. Bode, *Network Analysis and Feedback Amplifier Design,* pp. #137-169, D. Van Nostrand Company, Inc., New York, N.Y., **1945.**

4 W. **R.** Evans, *Control System Dynamics,* pp. #96-121, McGraw-Hill Book Company, Inc., New York, N. Y., **1954.**

5 W. R. Evans, "Control System Synthesis by Root-Locus Method", *Trans. AIEE,* Vol. **69,** Part I, pp. #66-69, **1950.**

6 **E. A.** Guillemin, *The Mathematics* of *Circuit Analysis,* pp. #395-409, John Wiley & Sons, Inc., New York, N.Y., **1949.**

CHAPTER 5

GAIN DETERMINATION*

5-1 PERFORMANCE CRITERIA AND DEFINITIONS[3,6]

5-1.1 GENERAL

Performance criteria are tests or rules by which one can determine, from the system parameters, whether or not the system has certain particular performance characteristics. An important parameter used in the design of servo systems is the gain of the system. The relations between gain and performance specifications are usually expressed in terms of performance constants, such as the velocity constant, the acceleration constant, and the torque'constant. Definitions of several imporant parameters and constants are given below.

5-1.2 GAIN

If $G(s)$ is the transfer function of a component, the gain K of the component is

$$K = \lim_{s \to 0} \; s^{\pm n} \, G(s) \tag{5-1}$$

where n is the order of the pole or zero of $G(s)$ at the origin. The plus sign is used for a simple or multiple pole at the origin, and the minus sign is used for a simple or multiple *zero* at the origin. The gain K is easy to identify if $G(s)$ is written in the form

$$G(s) = K \, s^{\pm n} \left[\frac{a_k s^k + a_{k-1} s^{k-1} + \ldots + 1}{b_j s^j + b_{j-1} s^{j-1} + \ldots + 1} \right] \tag{5-2}$$

5-1.3 VELOCITY CONSTANT

The velocity constant of a system is a measure of the steady-state error if the input to the system is a constant velocity. The velocity constant is defined by the relation

$$K_v = \left(\frac{E_{ss}}{\omega_i} \right)^{-1} \tag{5-3}$$

where

ω_i = constant input velocity

E_{ss} = steady-state error

For a single-loop unity-feedback system (Fig. 5-l) , the velocity constant is

$$K_v = \lim_{s \to 0} \left[s \frac{C(s)}{E(s)} \right] = \lim_{s \to 0} \left[s \, G(s) \right] \tag{5-4}$$

An analysis of **Eq. (5-4)** shows that the velocity constant of a single-loop unity-feedbitch system is finite and nonzero only if the open-loop transfer function $C(s)/E(s)$ has exactly one single-order pole at the origin (one integration).

5-1.4 ACCELERATION CONSTANT

The acceleration constant of a system is defined by the relation

$$K_a = \left(\frac{E_{ss}}{\alpha_i} \right)^{-1} \tag{5-5}$$

where

α_i = constant input acceleration

E_{ss} = steady-state error

*By L. A. Gould

For a single-loop unity-feedback system (Fig. 5-1), the acceleration constant is

$$K_A = \lim_{s \to 0} \left[s^2 \frac{C(s)}{E(s)} \right] = \lim_{s \to 0} [s^2 G(s)] \quad (5\text{-}6)$$

An analysis of Eq. (5-6) shows that the acceleration constant of a single-loop unity-feedback system is finite and nonzero only if the open-loop transfer function $C(s)/E(s)$ has exactly one double-order pole at the origin (two integrations).

5-1.5 TORQUE CONSTANT

The torque constant of a system is defined by the relation

$$K_T = \left(\frac{E_{ss}}{T_L} \right)^{-1} \quad (5\text{-}7)$$

where

T_L = constant load torque

E_{ss} = steady-state error

5-1.6 STATIC ACCURACY

The static accuracy of a ***linear*** system is measured by the steady-state error that is developed for a specified steady-state input or disturbance. A steady-state input in this context means that the input is a constant position, a constant velocity, a constant acceleration, etc. A steady-state disturbance means that the disturbance is a constant. The performance constants defined previously will uniquely determine these steady-state errors, as may be seen from the definitions of the constants.

5-1.7 BANDWIDTH

In general, the bandwidth of a servo system refers to a frequency interval between 0 and some upper frequency. There is no universally accepted definition of the upper frequency.

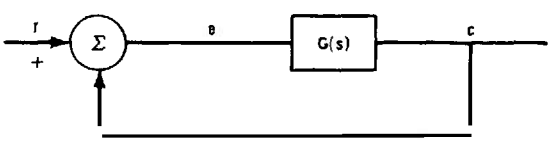


Fig. 5-1 Single-loop unity-feedback system.

Several commonly used upper-frequency values for unity-feedback systems (Fig. 5-1) are given below:

(a) ω_R, resonant frequency —frequency at which the closed-loop frequency response $C(j\omega)/R(j\omega)$ has its peak magnitude M_p (Fig. 5-2).

(b) ω_a —frequency at which the magnitude of the closed-loop frequency response $C(j\omega)/R(j\omega)$ is unity (Fig. 5-2).

(c) ω_b — frequency at which the magnitude of the closed-loop frequency response $C(j\omega)/R(j\omega)$ is 0.707 (Fig. 5-2).

(d) ω_p —frequency at which the phase of the closed-loop frequency response is —90° (Fig. 5-3).

(e) ω_e —frequency at which the magnitude of the error-to-input frequency response $E(j\omega)/R(j\omega)$ is 0.1 (Fig. 5-4).

(f) ω_{cm}, magnitude crossover frequency — frequency at which the magnitude of the open-loop frequency response $C(j\omega)/E(j\omega)$ is unity (Fig. 5-5).

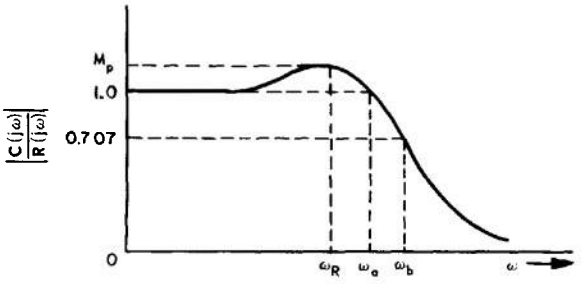


Fig. 5-2 Bandwidth measures from magnitude of closed-loop frequency response $C(j\omega)/R(j\omega)$.

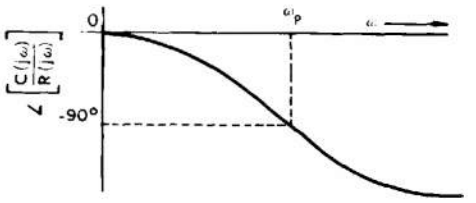


Fig. 5-3 Bandwidth measure from phase of closed-loop frequency response $C(j\omega)/R(j\omega)$.

(g) ω_c, asymptote crossover frequency — frequency at which the −10 dg/dec asymptote of the open-loop frequency response $C(j\omega)/E(j\omega)$ crosses 0 dg (Fig. 5-5; see Par. 5-3 for terminology).

5-1.8 PEAK MAGNITUDE

The peak magnitude M, is defined as the maximum value of the magnitude of the closed-loop frequency response $C(j\omega)/R(j\omega)$ or the magnitude of the resonant peak of the response (see Fig. 5-2).

The value of M_p is used as a measure of the degree of stability, and design in the frequency domain usually involves adjusting a gain K so as to satisfy a specified value of M_p. Large values of M_p are indicative of highly oscillatory behavior, whereas values of M_p less than unity are indicative of heavily damped behavior. In practice, M_p usually lies between 1.3 and 1.6; that is, the range of M_p is usually specified as follows:

$$1.3 < M_p < 1.6 \qquad (5\text{-}8)$$

or

$$1 \text{ dg} < 10 \log_{10} M_p < 2 \text{ dg} \qquad (5\text{-}9)$$

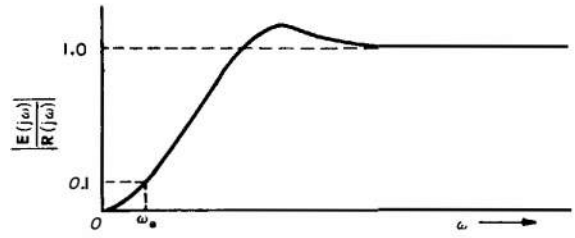


Fig. 5-4 Bandwidth measure from magnitude of error-to-input frequency response $E(j\omega)/R(j\omega)$.

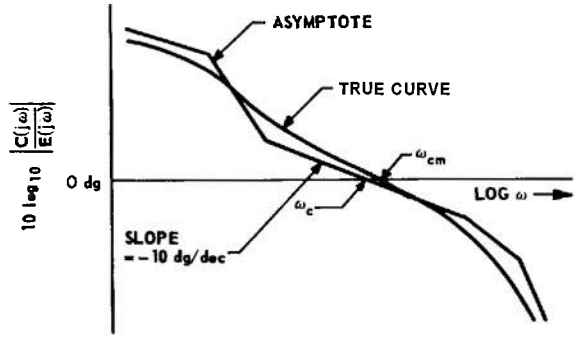


Fig. 5-5 Bandwidth measures from open-loop frequency response.

5-2 POLAR-PLANE REPRESENTATION[1,2,9,12]

5-2.1 GENERAL

A polar-plane representation of the open-loop frequency response $C(j\omega)/E(j\omega)$ is often used in the process of carrying out a design in the frequency domain. A plot of $C(j\omega)/E(j\omega)$ in polar coordinates makes it easier to apply the Nyquist stability criterion to determine gain setting ranges for stable operation. In addition, the determination of gain for a specified M_p value involves only a simple graphical construction on the polar plane (see Par. 5-4). Plots of both the direct function and its inverse are used, wherein only positive frequencies are usually considered.

5-2.2 DIRECT POLAR PLANE

A direct polar-plane plot of the function $G(j\omega)$ is constructed by drawing a curve through the points whose polar coordinates at each frequency are the magnitude of $G(j\omega)$ and the phase angle of $G(j\omega)$ at that frequency, where the phase angle of $G(j\omega)$ is the phase of $c(t)$ minus the phase of $e(t)$ when $e(t)$ and $c(t)$ are sinusoids. Positive angles are plotted in a counterclockwise direction. Increasing the gain associated with $G(j\omega)$ expands the polar locus in a radial direction. If $G(j\omega)$ is cascaded with another transfer function, the resulting polar coordinates of the combination are obtained- at

each frequency by: (1) multiplying the magnitude of $G(j\omega)$ by the magnitude of the cascaded function to give the magnitude of the combination; and (2) adding the phase angle of the cascaded function to the phase angle of $G(j\omega)$ to give the phase angle of the combination.

5-2.3 INVERSE POLAR PLANE

An inverse polar-plane plot of a function is a plot of the reciprocal of the function on the polar plane. The reciprocal or inverse of $G(j\omega)$ is written as follows:

$$G^{-1}(j\omega) = \frac{1}{G(j\omega)} \qquad (5\text{-}10)$$

The polar coordinates of $G^{-1}(j\omega)$ at each frequency are given by: (1) the reciprocal of the magnitude of $G(j\omega)$; and (2) the negative of the phase angle of $G(j\omega)$. Increasing the gain of $G(j\omega)$ shrinks the inverse locus in the radial direction. If $G(j\omega)$ is cascaded with another transfer function, the resulting polar coordinates of the inverse of the combination are obtained at each frequency by: (1) multiplying the magnitude of the inverse $G(j\omega)$ locus by the inverse of the magnitude of the cascaded function to give the magnitude of the inverse of the combination; and (2) adding the negative of the phase angle of the cascaded function to the phase of the inverse $G(j\omega)$ locus to give the phase angle of the inverse of the combination.

Example. Plots of the direct function

$$G(j\omega) = \frac{1}{j\omega\,[(j\omega)^2 + 0.6j\omega + 1]} \qquad (5\text{-}11)$$

and a multiple of its inverse $3G^{-1}(j\omega)$ appear in Fig. 5-6.

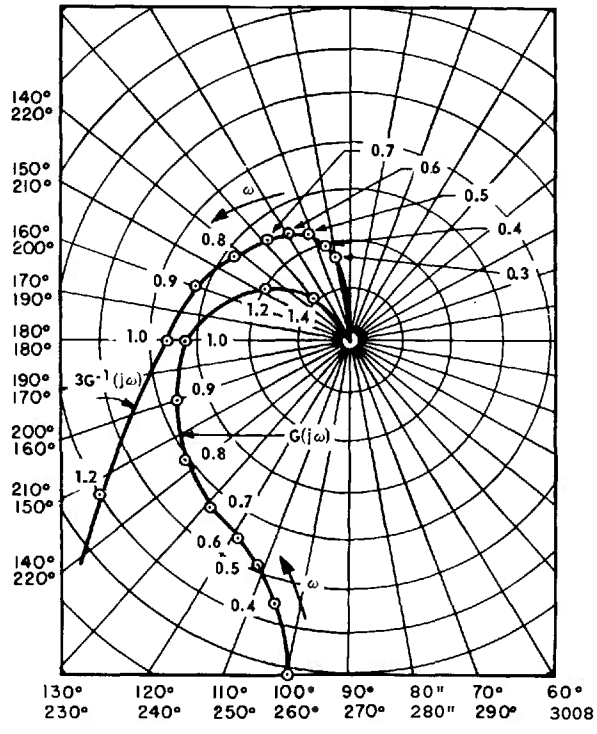


Fig. 5-6 Direct and inverse polar plots of

$$G(j\omega) = \frac{1}{j\omega[(j\omega)^2 + 0.6j\omega + 1]}.$$

5-3 EXACT AND ASYMPTOTIC-LOGARITHMIC REPRESENTATIONS [5,9,13]

5-3.1 GENERAL

The logarithmic method of representing a function is a more convenient way to present frequency-response information than the polar-plane method. The advantage of the logarithmic procedure is that magnitude multiplication for cascaded functions reduces to the simple addition of logarithms. Furthermore, the magnitudes of the first- and second-order factors of transfer functions can readily be approximated by straight-line asymptotes when the functions are plotted to a logarithmic scale. Such asymptotic approximations reduce the time taken up by calculation and, in addition, enable the designer to make a rough estimate of system performance, when this is necessary.

5-3.2 SEPARATE MAGNITUDE AND PHASE PLOTS

The separate magnitude and phase-angle plots for a transfer function $G(j\omega)$ are respectively: (1) plots of $10 \log_{10} |G(j\omega)|$ versus $\log \omega$; and (2) plots of Ang $G(j\omega)$ versus $\log \omega$. The unit of logarithmic magnitude used in these plots is called the *decilog*, abbreviated dg. The *magnitude of a number N in decilogs is* $10 \log_{10} N$. For convenience in plotting, semilog graph paper is generally used. To plot a transfer function $G(j\omega)$ that is already in factored form, several aids (to be discussed below) are available which simplify the procedure. Before discussing these aids, however, it is helpful to point out the general types of factors that may appear in any rational algebraic function. Consider a function $G(j\omega)$ whose factored form can be written as follows:

$$G(j\omega) = K\,(j\omega)^{\pm n}\,\frac{(T_1 j\omega + 1)\left[\left(j\,\frac{\omega}{\omega_{n_1}}\right)^2 + 2\zeta_1 j\,\frac{\omega}{\omega_{n_1}} + 1\right]\ldots}{(T_2 j\omega + 1)\left[\left(j\,\frac{\omega}{\omega_{n_2}}\right)^2 + 2\zeta_2 j\,\frac{\omega}{\omega_{n_2}} + 1\right]\ldots} \qquad (5\text{-}12)$$

Only three general types of algebraic factors appear in Eq. (5-12).

The three factor types, which may occur in any rational function, are the following:

$(j\omega)^{\pm n}$ (differentiation or integration) (5-13)

$(Tj\omega + 1)$ (first order) (5-14)

$\left[j\left(\frac{\omega}{\omega_n}\right)^2 + 2\zeta j\,\frac{\omega}{\omega_n} + 1\right]$ (second order) (5-15)

5-3.3 MAGNITUDE CURVES

The magnitude curve of the quantity $(j\omega)^{\pm n}$ is a straight line passing through 0 dg at $\omega = 1$ with a slope equal to $\pm 10\ n$ dg/decade.

The magnitude of the first-order factor $(Tj\omega + 1)^{\pm 1}$ can be approximated by two straight lines. For $T\omega << 1$, the asymptote is the 0-dg line. For $T\omega >> 1$, the asymptote is a line with a slope of ± 10 dg/decade that crosses Odg at $T\omega = 1$. The frequency $\omega_b = 1/T$ is called the *break* frequency of the factor. The true magnitude curve can be obtained from the asymptotes by applying the two rules-of-thumb:

(a) At tht break frequency, the true curve is 1.5 dg above (or below) the asymptotes.

(b) At an octave above and below the break frequency, the true curve lies 0.5 dg above (or below) the asymptotes.

The asymptotes and the true magnitude curves for a first-order factor are shown in Fig. 5-7. Note in this figure that the magnitude curve of $(Tj\omega = 1)^{-1}$ is the mirror image of the magnitude curve of $(Tj\omega + 1)$ about the 0-dg line.

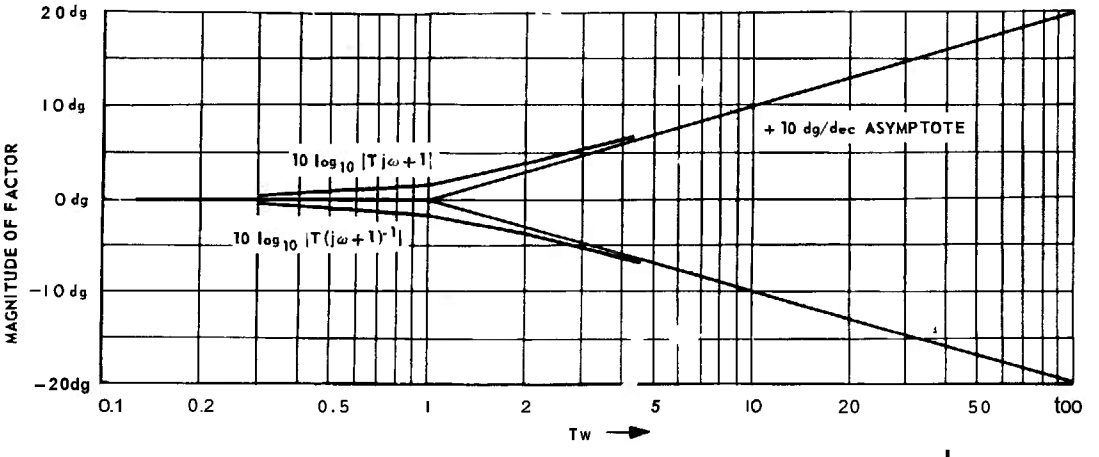


Fig. 5-7 Asymptotes and true magnitude curves for the first-order factor $(Tj\omega + 1)^{+1}$.

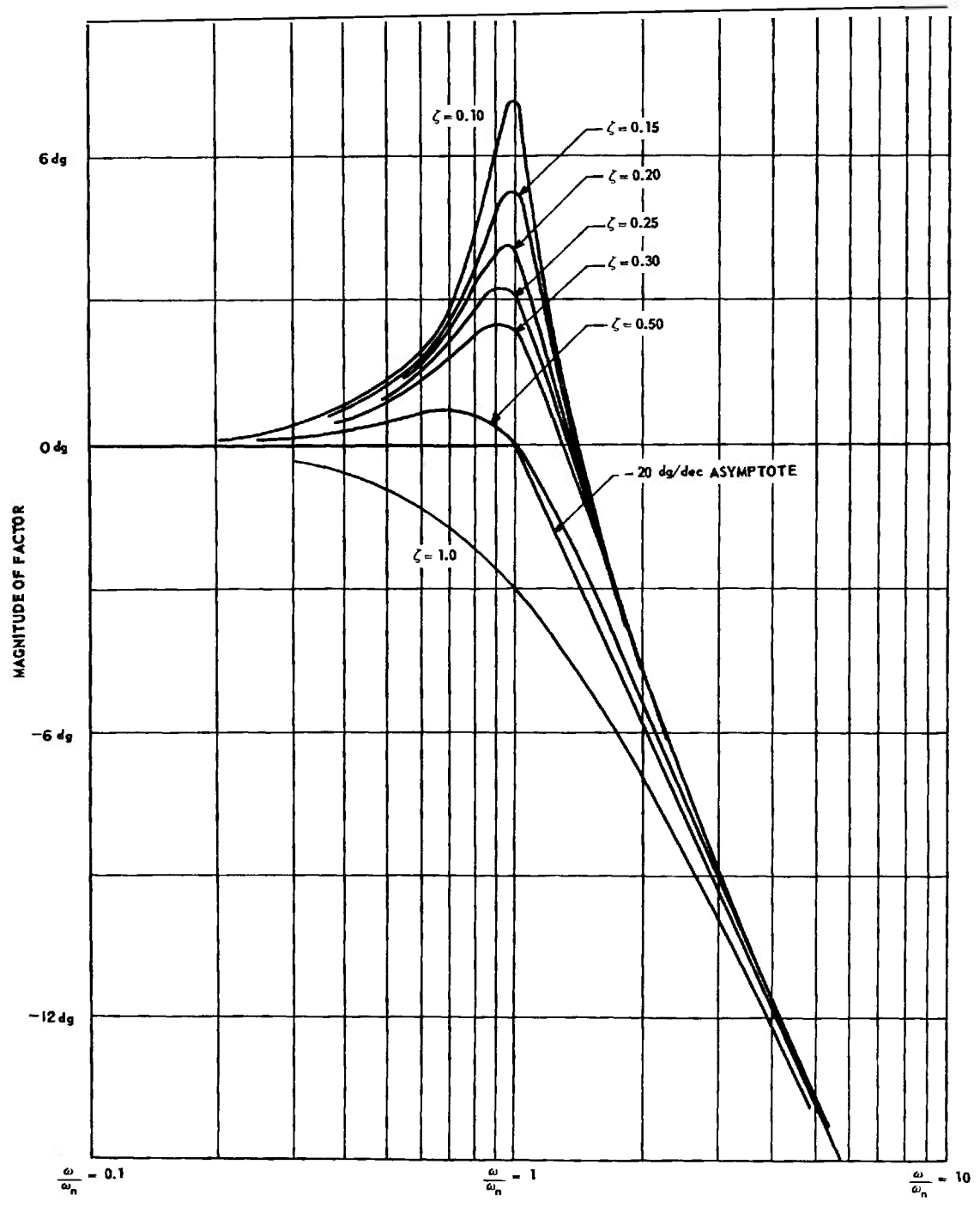


Fig. 5-8 Asymptotes and true curves for the second-order factor $\left[\left(i\frac{\omega}{\omega_n}\right)^2 + 2\zeta i\frac{\omega}{\omega_n} + 1\right]^{-1}$.

The magnitude of the second-order factor

$$\left[\left(j\frac{\omega}{\omega_n}\right)^2+2\zeta j\frac{\omega}{\omega_n}+1\right]^{\pm 1}$$

can be approximated by two straight-line asymptotes. For $\omega<<\omega_n$, the asymptote is the 0-dg line. For $\omega>>\omega_n$, the asymptote is a straight line with a slope of t2 0 dg/decade crossing the 0-dg line at the break frequency, $\omega_b = \omega_n$. A set of second-order magnitude curves is shown in Fig. 5-8 for different values of the damping ratio ζ. Not that the approximation is best for $\zeta = 0.5$.

5-3.4 PHASE-ANGLE CURVES

The phase angle of the factor $(j\omega)^{\pm n}$ is a constant equal to $\pm 90n°$. The phase-angle curves of the first-order factor $(Tj\omega+1)^{\pm 1}$ are shown in Fig. **5-9.** Note that each curve is symmetrical about the point on the curve at which $\omega = 1/T$. The phase-angle curves of the second-order factor

$$\left[\left(j\frac{\omega}{\omega_n}\right)^2+2\zeta j\frac{\omega}{\omega_n}+1\right]$$

are shown in Fig. 5-10 for different values of the damping ratio ζ.

To plot the separate magnitude and phase-angle curves for a factored transfer function, separate plots are first made of the magnitude and phase-angle curves of each factor. In doing this, care must be taken to distinguish between numerator and denominatur factors. Next, all the magnitudes are added at each frequency to obtain the composite magnitude curve. Similarly, all the phase angles are added at each frequency to obtain the composite phase-angle curve. It is important to note that the factors must be in the standard forms given in Eqs. (5-8), (5-9),

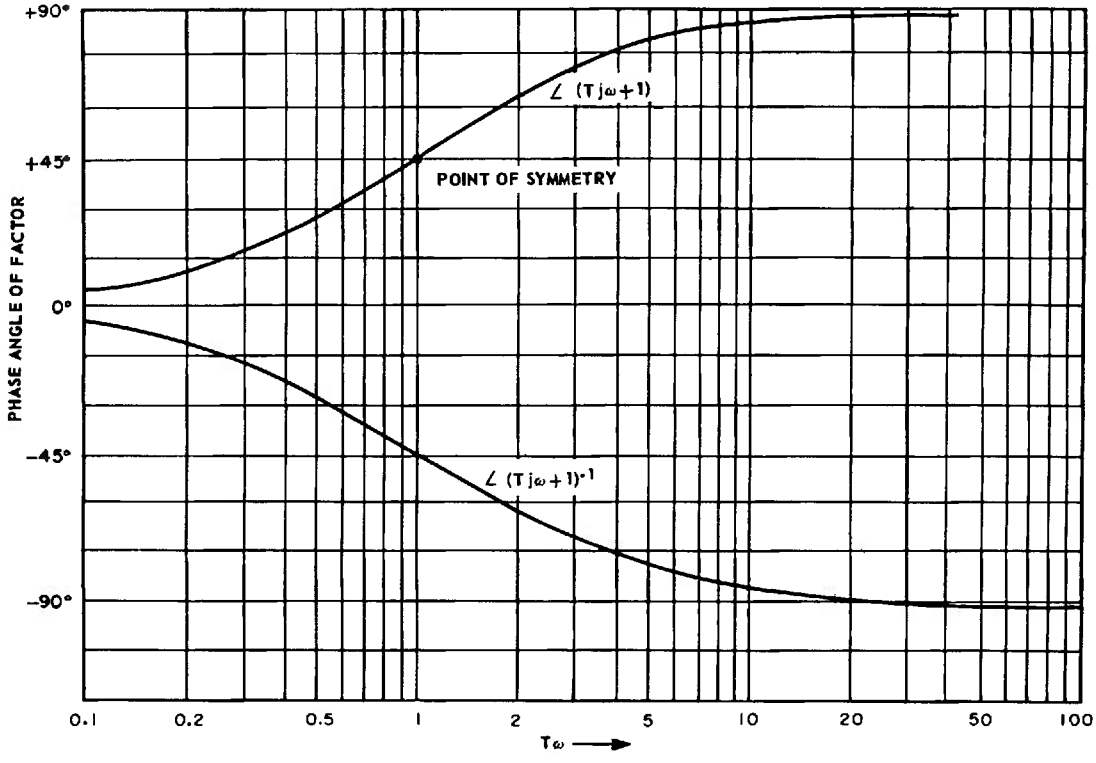


Fig. 5-9 Phase-angle curves for the first-order factor $(Tj\omega+1)^{\pm 1}$.

and (5-10) if the curves of Figs. 5-7 to 5-10 are to be used. The effect of the gain K can be incorporated by merely adding $10 \log_{10} K$ to the magnitude scale of the composite magnitude curve.

Example. The separate factors and composite curves for the function

$$G(j\omega) = K \frac{(0.2j\omega + 1)}{(j\omega)\left[\left(j\frac{\omega}{10}\right)^2 + 0.6\, j\frac{\omega}{10} + 1\right]} \tag{5-16}$$

where $K = 6.5$ are plotted in Figs. 5-11 and 5-12.

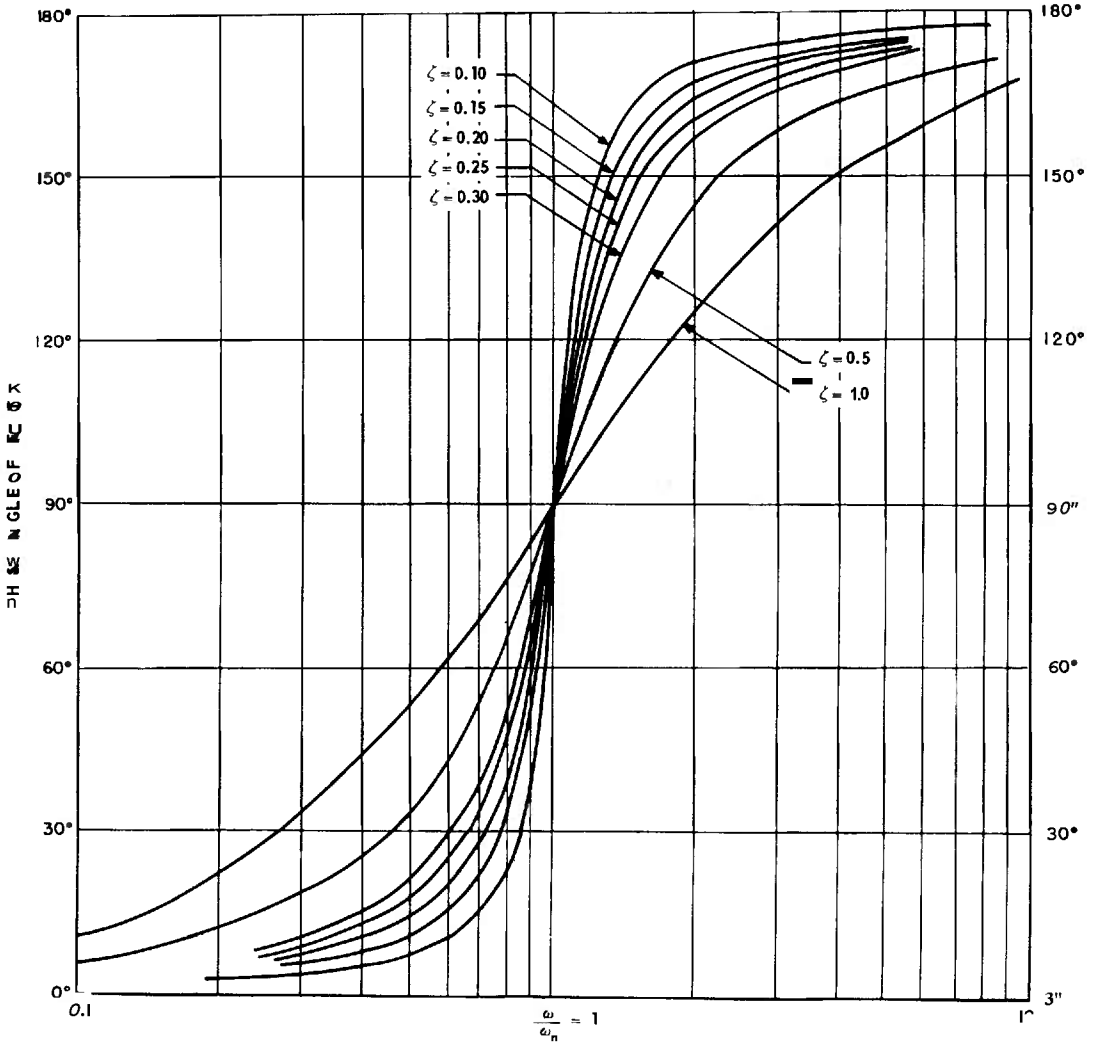


fig. 5-10 Phase-angle curves for the second-order factor $\left[\left(j\frac{\omega}{\omega_n}\right)^2 + 2\zeta j\frac{\omega}{\omega_n} + 1\right]$.

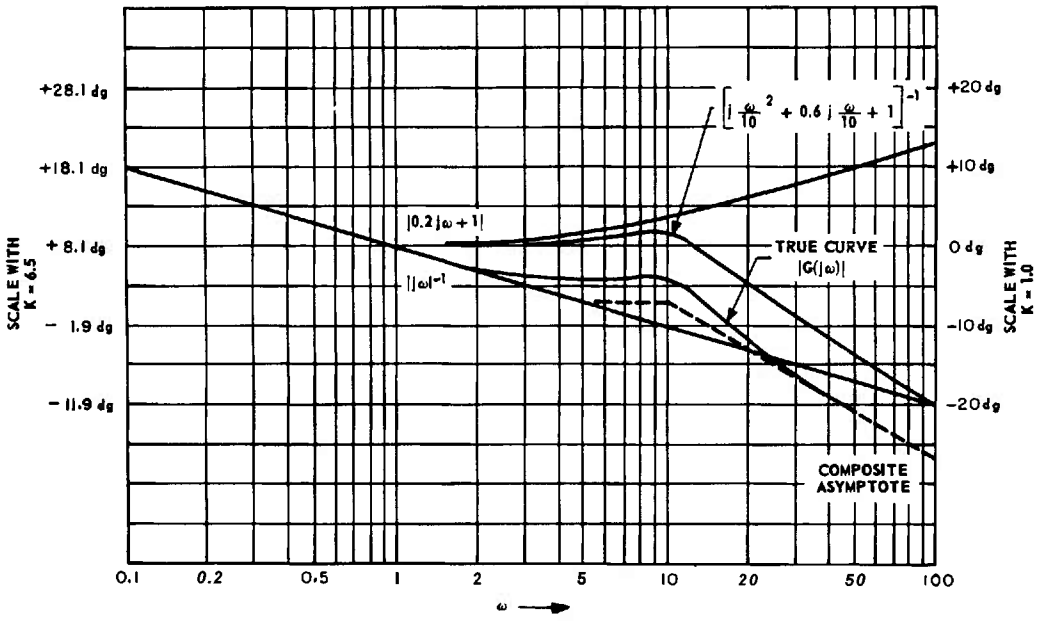


Fig. 5-77 Magnitude plots for $G(j\omega) = K\,\dfrac{[0.2\,j\omega + 1]}{j\omega\left[\left(j\dfrac{\omega}{10}\right)^2 + 0.6\,j\dfrac{\omega}{10} + 1\right]}$

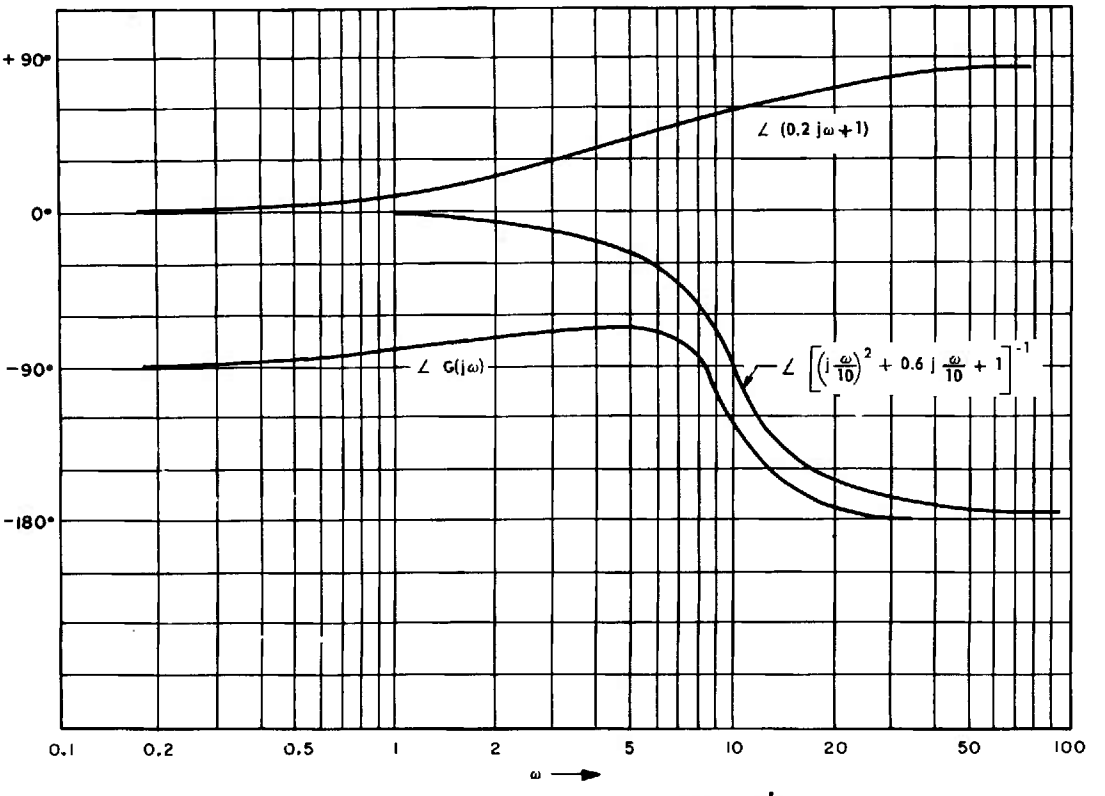


Fig. 5-72 Angle plots for $G(j\omega) = K\,\dfrac{(0.2j\omega + 1)}{j\omega\left[\left(j\dfrac{\omega}{10}\right)^2 + 0.6\,j\dfrac{\omega}{10} + 1\right]}$

5-3.5 GAIN-PHASE PLANE

To facilitate design, a third method for representing frequency functions may be used. In this method, the magnitude and phase angle of a frequency function are plotted on a coordinate system called the *gain-phase plane.* The magnitude is plotted to a logarithmic scale (in decilogs) and the phase angle is plotted to a linear scale. Frequency is the parameter for the gain-phase plot. The gain-phase plot can be determined directly from the frequency function by calculating the magnitude (in dg) and phase angle of the function at various frequencies. Alternatively, the gain-phase plot can be determined through the intermediate use of the separate magnitude and phase-angle plots when the function to be plotted is in factored form. The gain-phase plot is most useful for determining the closed-loop response of a system from the open-loop response (see Pars. 5-4 and 5-5).

Example. The function

$$G(j\omega) = 6.5\frac{(0.2j\omega + 1)}{(j\omega)\left[\left(j\frac{\omega}{10}\right)^2 + 0.6\,j\frac{\omega}{10} + 1\right]} \qquad (5\text{-}17)$$

is plotted on the gain-phase plane in Fig. 5-13.

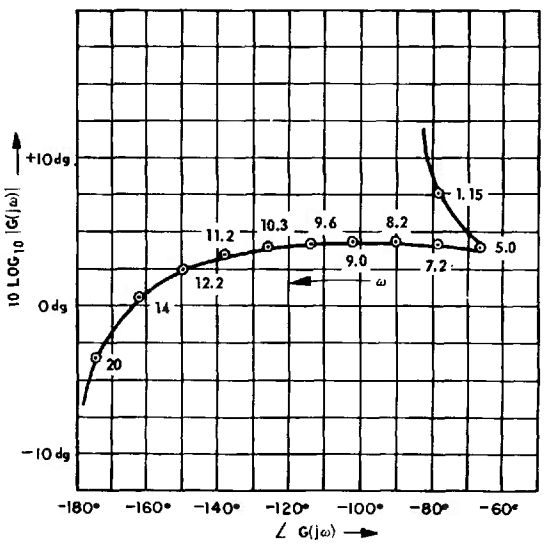


Fig. 5-73 Gain-phase plot of $G(j\omega) = 6.5\dfrac{(0.2j\omega + 1)}{j\omega\left[\left(j\frac{\omega}{10}\right)^2 + 0.6\,j\frac{\omega}{10} + 1\right]}$

5-4 CLOSED-LOOP RESPONSE DETERMINATION[4,5,9,12,13,14]

5-4.1 GENERAL

The relations that exist between the closed- and open-loop responses of a unity-feedback system can be obtained by considering the diagram in Fig. 5-1. In this diagram, it is clear that the open-loop responses is given by

$$\frac{C(j\omega)}{E(j\omega)} = G(j\omega) \qquad (5\text{-}18)$$

The closed-loop response function $W(j\omega)$ is defined as follows:

$$\frac{C(j\omega)}{R(j\omega)} \triangleq W(j\omega) \qquad (5\text{-}19)$$

Now, $E(j\omega) = R(j\omega) - C(j\omega)$. Substituting this expression for $E(j\omega)$ into Eq. (5-18), rearranging terms, and using the definition in Eq. (5-19), it is found that

$$W(j\omega) = \frac{G(j\omega)}{1 + G(j\omega)} \qquad (5\text{-}20)$$

The transformation from the G plane to the W plane defined by Eq. (5-20) is used to determine the closed-loop response $W(j\omega)$ from the corresponding open-loop response $G(j\omega)$. Although a direct calculation of the W function is possible, this is often avoided

because it usually proves to be tedious. Instead, various aids for performing the G-to-W plane transformation are used. These are presented next.

5-4.2 POLAR-PLANE TECHNIQUE

In the polar-plane technique, a "vector" construction on the G plane is used to determine both the magnitude and phase angle of W. As an illustration, consider the G function sketched in Fig. **5-14.** In this figure, once the $-1 + j0$ point is located, the following "vector" relations hold :

$$\overline{OB} = G \tag{5-21}$$

$$\overline{OA} = -1 \tag{5-22}$$

$$\overline{OB} - \overline{OA} = \overline{AB} = 1 + G \tag{5-23}$$

Then, the closed-loop response at each frequency can be determined from the construction of Fig. **5-14** as follows:†

$$|W(j\omega)| = \frac{|\overline{OB}|}{|\overline{AB}|} \tag{5-24}$$

$$\angle\, W(j\omega) = \angle\, ABO = \phi \tag{5-25}$$

†Symbol $\angle$ denotes "angle"

The inverse G-plane construction for the closed-loop response is shown in Fig. **5-15.** In this figure, the following "vector" relations hold :

$$\overline{OB} = G^{-1} \tag{5-26}$$

$$\overline{OA} = -1 \tag{5-27}$$

$$\overline{AO} + \overline{OB} = \overline{AB} = 1 + G^{-1} \tag{5-28}$$

The closed-loop response can be determined from the following relations :

$$|W(j\omega)| = \frac{|\overline{AO}|}{|\overline{AB}|} \tag{5-29}$$

$$\angle\, W(j\omega) = \angle\, OAB = \phi \tag{5-30}$$

To avoid "vector" constructions, constant magnitude and constant phase-angle contours that correspond to the G-to-W transformation of Eq. **(5-20)** are often used. The following definitions apply :

$$M = |W(j\omega)| \tag{5-31}$$

$$\phi = \angle\, W(j\omega) \tag{5-32}$$

$$N = \tan \phi \tag{5-33}$$

The transformation given in Eq. **(5-20)** can be used to map contours of constant M and constant N onto the G plane. The M contours appear as a set of bipolar circles as

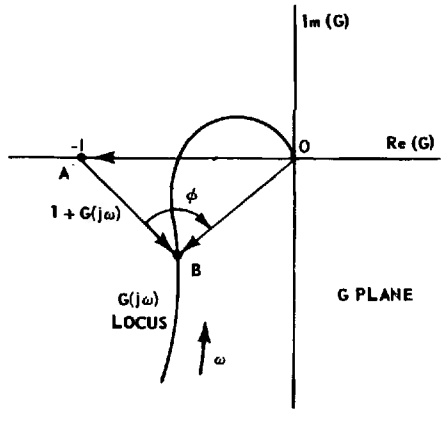


Fig. 5-14 Closed-loop response construction on the G plane.

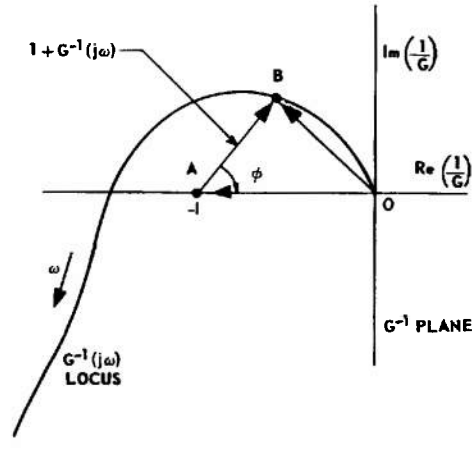


Fig. 5-15 Closed-loop response construction on the G^{-1} plane.

shown in Fig. **5-16.** The contours of constant N or ϕ are shown in Fig. **5-17.**

To represent the constant M and ϕ contours in the G^{-1} plane, **Eq.** ***(5-20)*** is used. The M contours are a set of concentric circles and the ϕ contours are a set of straight lines. The M and ϕ contours for the G^{-1} plane are shown in Fig. **5-18.**

The properties of the M and ϕ contours are listed in Table 5-1.

The M and ϕ contours are the lines or curves of constant M and constant ϕ as they appear in the G or G^{-1} planes. By constructing a chart of M and ϕ circles for the G or G^{-1} planes, one has the coordinate system of the W plane represented by circles and lines in the G or G' plane. The closed-loop magnitude M and phase angle ϕ can be obtained directly from the G function by constructing the G function (or the G^{-1} function) on a chart of constant M and + contours. At each frequency, the value of the M contour that intersects the G (or G^{-1}) function is the value of the magnitude of the closed-loop response W. Similarly, the value of the ϕ contour that intersects the G (or G^{-1}) function at a given frequency is the value of the phase angle of the closed-loop response W.

The M and ϕ contours aid greatly in performing the transformation from open- to closed-loop frequency response and are used to facilitate the design of a system when the shape of the G function is to be altered so as to improve performance.

TABLE 5-1 PROPERTIES OF M AND ϕ CONTOURS

G Plane	G^{-1} Plane
M contours	M contours
$y = Im(G)$	$y = Im(G^{-1})$
$x = Re(G)$	$x = Re(G^{-1})$
$y^2 + \left(x + \frac{M^2}{M^2-1}\right)^2 = \frac{M^2}{(M^2-1)^2}$	$y^2 + (1+x)^2 = \frac{1}{M^2}$
center: $\left(-\frac{M^2}{M^2-1}, 0\right)$	center: $(-1, 0)$
radius: $\left\lvert\frac{M}{M^2-1}\right\rvert$	radius: $\frac{1}{M}$
intercept nearest origin : $-\frac{M}{M+1}$	intercept nearest origin : $\frac{1-M}{M}$
N contours	N contours
$N = \tan\phi$	$N = \tan\phi$
$(x+0.5)^2 + \left(y - \frac{1}{2N}\right)^2 = \frac{1}{4}\left(\frac{N^2+1}{N^2}\right)$	$y^2 + Nx + N = 0$
center: $\left(-\frac{1}{2}, \frac{1}{2N}\right)$	Note; N contours are a family of radial lines emanating from the center of the M circles
radius: $\frac{1}{2}\frac{\sqrt{N^2+1}}{N}$	

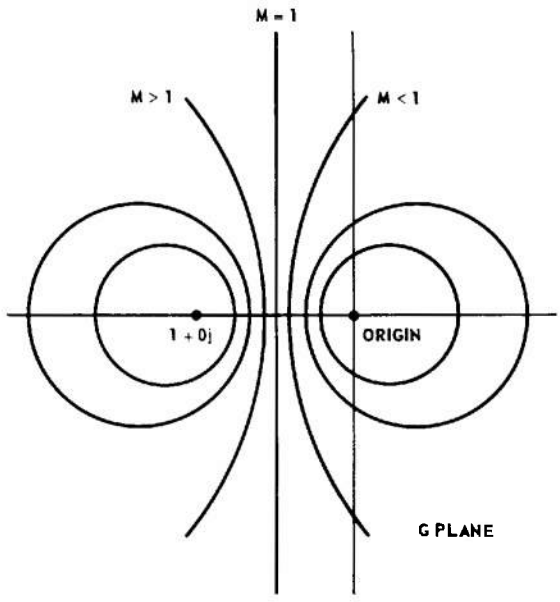


fig. 5-16 Contours of constant M in the G plane.

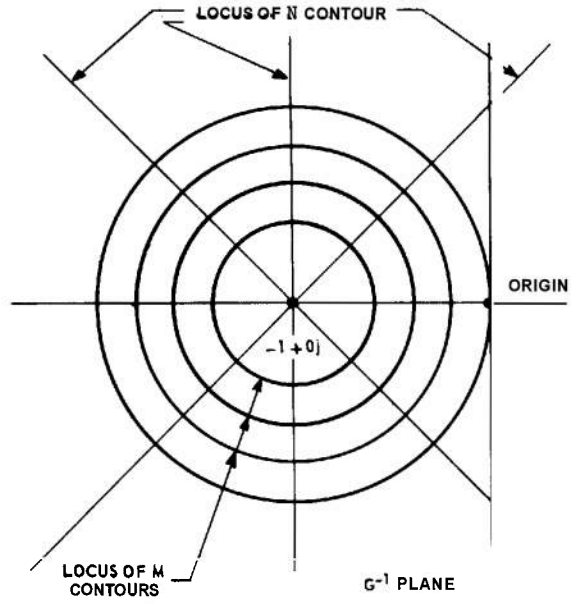


fig. 5-18 Contours of constant M and constant ϕ in the G^{-1} plane.

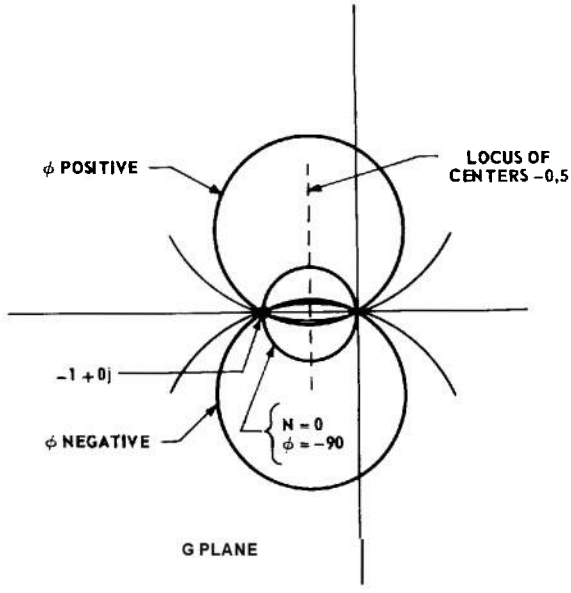


Fig. 5-17 Contours of constant phase in the G plane.

5-4.3 GAIN-PHASE PLANE TECHNIQUE (NICHOLS CHART)

Since constructions on the gain-phase plane involving cascaded functions and gain alterations are usually simpler than similar constructions on the polar plane, a chart of constant M and ϕ contours has been constructed for the gain-phase plane. This chart is called the Nichols chart and is shown in Figs. 5-19 and 5-20. Figure 5-19 presents a

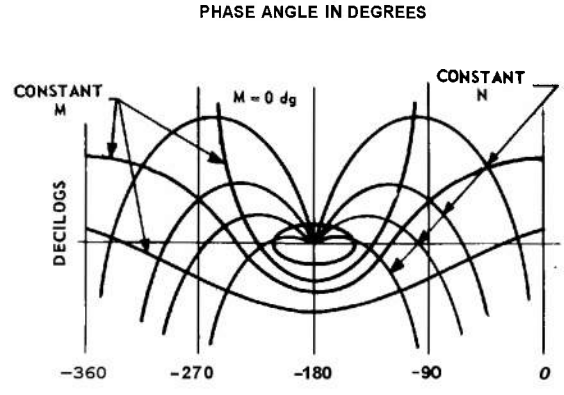


Fig. 5-19 Chart showing symmetry of M-N contours about phase of 180 degrees (Nichols Chart).

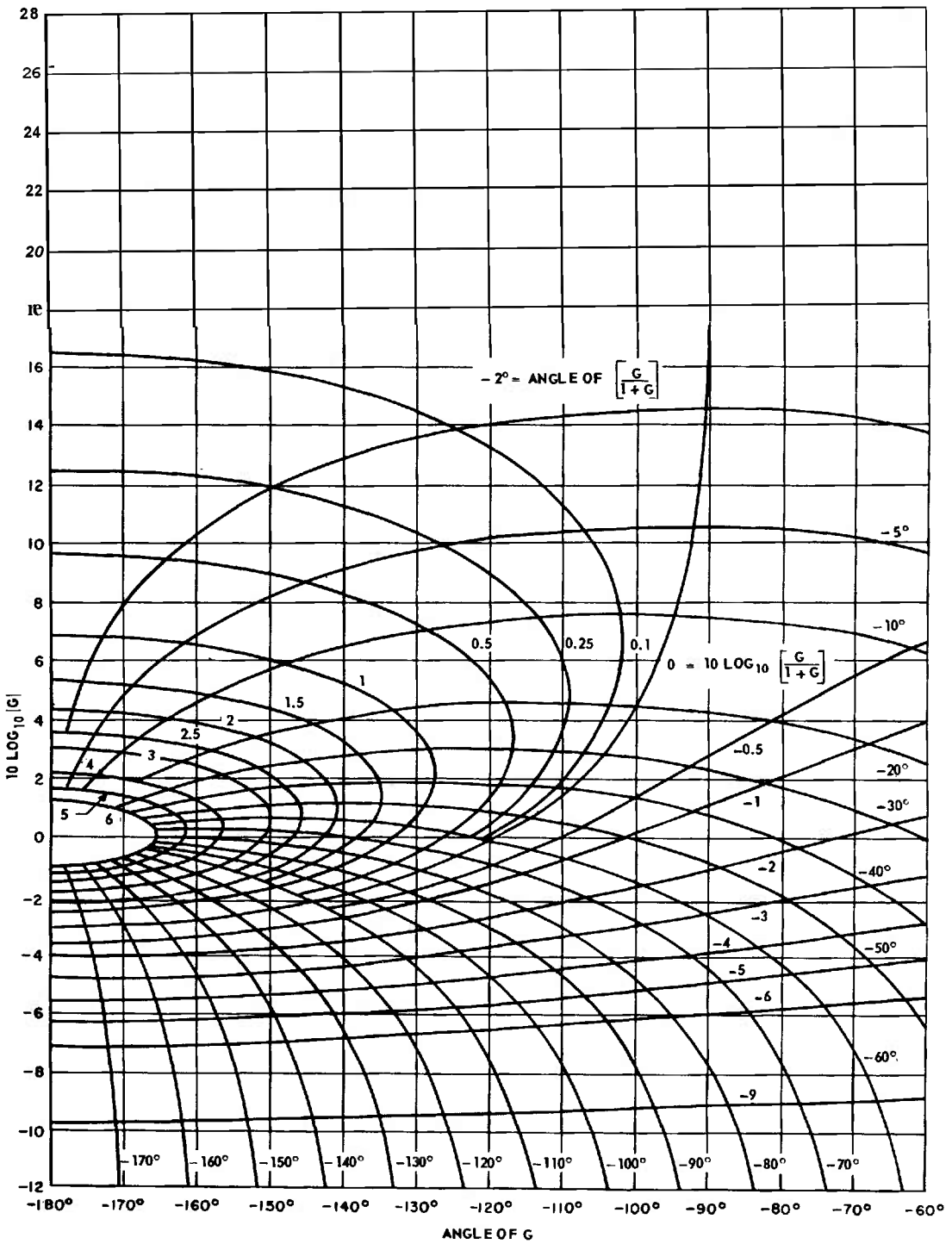


Fig. 5-20 Nichols chart.

large-scale view of the Nichols chart and Fig. 5-20 presents only that part of the Nichols chart that is most useful for design purposes.

The Nichols chart is used in the same way that the M-N contours are used on the polar plane. The G function is plotted on the Nichols chart. The value of the M contour that intersects the G function at a given frequency is the value of the magnitude of the closed-loop response W at that frequency. Similarly, the intersection of the G function with the ϕ contours determines the phase angle of the closed-loop response W as a function of frequency.

54.4 NONUNITY-FEEDBACK SYSTEMS

If the closed-loop response of a nonunity-feedback system is sought, a slight modification of the procedure used for the unity-feedback system will enable the designer to use the Nichols chart and the polar M-N contours as well.

The closed-loop response of the nonunity-feedback system (Fig. 5-21) can be written as follows:

$$\frac{C(j\omega)}{R(j\omega)} = \frac{1}{H(j\omega)} \left[\frac{G(j\omega)\,H(j\omega)}{1 + G(j\omega)\,H(j\omega)} \right] \tag{5-34}$$

Since the bracketed portion of the right-hand side of Eq. (5-34) has the same form as the right-hand side of Eq. (5-20), the Nichols chart (or the polar M-N contours) can be used to find $GH/(1+\mathrm{GH})$ from a plot of $G(j\omega)\,H(j\omega)$. The closed-loop response $C(j\omega)/R(j\omega)$ can then be found by multiplying $GH/(1+\mathrm{GH})$ by H^{-1} at each frequency.

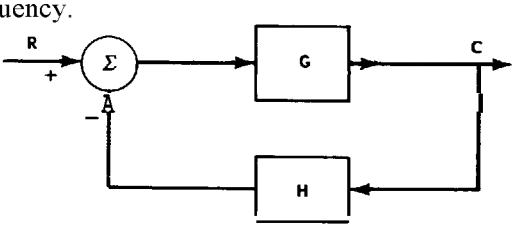


Fig. 5-27 Nonunity-feedback system.

5-5 SETTING THE GAIN FOR A SPECIFIED M_p [4,5,9,12,13,14]

5-5.1 GENERAL

A primary problem encountered in servo system design is the determination of the loop gain K required to produce a specified degree of stability. For a unity-feedback system (Fig. 5-1), the stability of the system is determined by the location of the $G(j\omega)$ locus with respect to the point $-1+j0$ (see Nyquist criterion, Par. 4-3). For a nonunity-feedback system (Fig. 5-21), however, the stability of the system is determined by the location of the $G(j\omega)\,H(j\omega)$ locus with respect to the point $-1+j0$. One analytical approach can serve for both types of systems if it is noted that, by redrawing Fig. 5-21, the study of the stability of a nonunity-feedback system can be expressed in terms of the stability of a unity-feedback system cascaded with another transfer function (Fig. 5-22). Thus, the discussion of stability can be limited to unity-feedback systems.

A system is said to have a low degree of stability if the normal mode of response is highly oscillatory. Such a system is also said

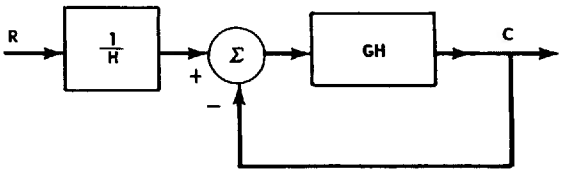


Fig. 5-22 System equivalent to the system of Fig. 5-21.

to have a **low** relative stability. The degree of stability of **a** stable unity-feedback system can be measured by the closeness of approach of the $G(j\omega)$ locus to the point $-1+j0$. An examination of the polar M contours (Figs. 5-16 and **5-18)** or the Nichols chart (Fig. 5-20) shows that, the larger the value of M, the more closely the $G(j\omega)$ contour approaches the point $-1+j0$. The peak magnitude of the closed-loop response $W(j\omega)$ is called M_p. By limiting this peak value, the degree of stability of a system can be maintained within a specified bound.

If $G(j\omega)$ is specified, except for a factor K, the degree of stability of the closed-loop response $W(j\omega)$ corresponding to this $G(j\omega)$ can be changed by adjusting the value of K. If the degree of stability as measured by M_p is specified, K is uniquely determined. The determination of K for a specified M_p is usually accomplished by a graphical construction in the polar or gain-phase plane.

5-5.2 POLAR-PLANE CONSTRUCTION

The polar-plane construction required to determine K [the gain of $G(j\omega)$] for a specified M_p is shown in Fig. **5-23** for the G/K plane and **Fig. 5-24** for the KG^{-1} plane.

The procedure used for a construction on the direct, or G/K, plane (Fig. **5-23)** is as follows:

(a) $\frac{G(j\omega)}{K}$ is plotted as a function of ω.

(b) **A** straight line is drawn from the origin, making an angle ψ with the real axis, where

$$\psi = \sin^{-1}\left(\frac{1}{M_p}\right)$$

This line is called the ψ line.

(c) **A** circle with center on the real axis is constructed, tangent to the ψ line and the $G(j\omega)/K$ locus.

(d) **A** line is drawn from the point of tangency of the ψ line with the circle (point b in Fig. **5-23)** normal to the real axis.

(e) The value of $Re(G/K)$ at the point of intersection of the normal with the real axis (point a in Fig. **5-23)** is the reciprocal of the gain K which must multiply $G(j\omega)/K$ to produce the specified M_p.

(f) The angulsr frequency at the point of tangency of the $G(j\omega)/K$ locus with the circle is the resonant frequency of the closed-loop system having the specified M_p (ω_R in Fig. **5-23).**

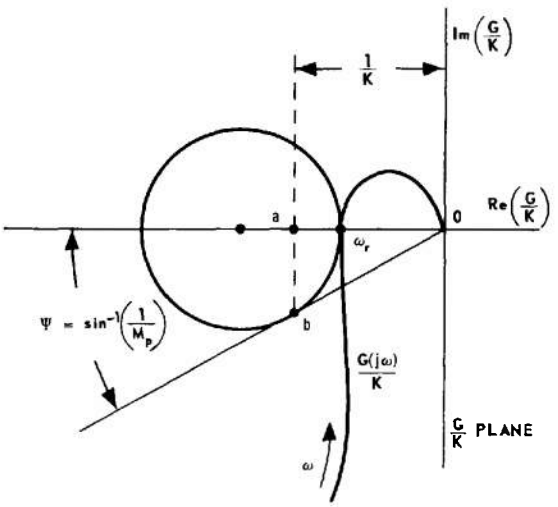


Fig. 5-23 Construction for gain determination on direct (G/K) plane.

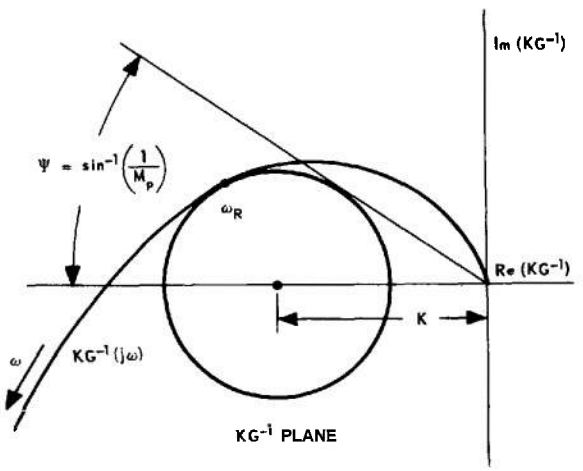


Fig. 5-24 Construction for gain determination on inverse (KG^{-1}) plane.

The procedure used for a construction on the inverse, or KG^{-1}, plane (Fig. **5-24**) is as follows:

(a) $KG^{-1}(j\omega)$ is plotted as a function of ω.

(b) A straight line (ψ line) is drawn from the origin, forming an angle ψ with the real axis, where

$$\psi = \sin^{-1}\left(\frac{1}{M_p}\right)$$

(c) A circle with center on the real axis is constructed, tangent to the KG^{-1} locus and the ψ line.

(d) The center of the circle is the point $-K+j0$. Thus, the coordinate of the center of the circle on the real axis is the value of K used to multiply $G(j\omega)$ to produce the desired M_p.

(e) The angular frequency at the point of tangency of the circle with the $KG^{-1}(j\omega)$ locus is the resonant frequency of the closed-loop system having the specified M_p (ω_R in Fig. **5-24**).

Example. A unity-feedback system has the following open-loop frequency response:

$$G(j\omega) = \frac{K}{j\omega[(j\omega)^2 + 0.6j\omega + 1]} \qquad (5\text{-}35)$$

Find K and ω_R for $M_p = 1.6$.

Solution.

(a) Direct-plane procedure:

(1) The $\frac{G(j\omega)}{K}$ locus is plotted (Fig. 5-25).

(2) The ψ line is drawn with $\psi = \sin^{-1}\left(\frac{1}{1.6}\right) = 38.7°$.

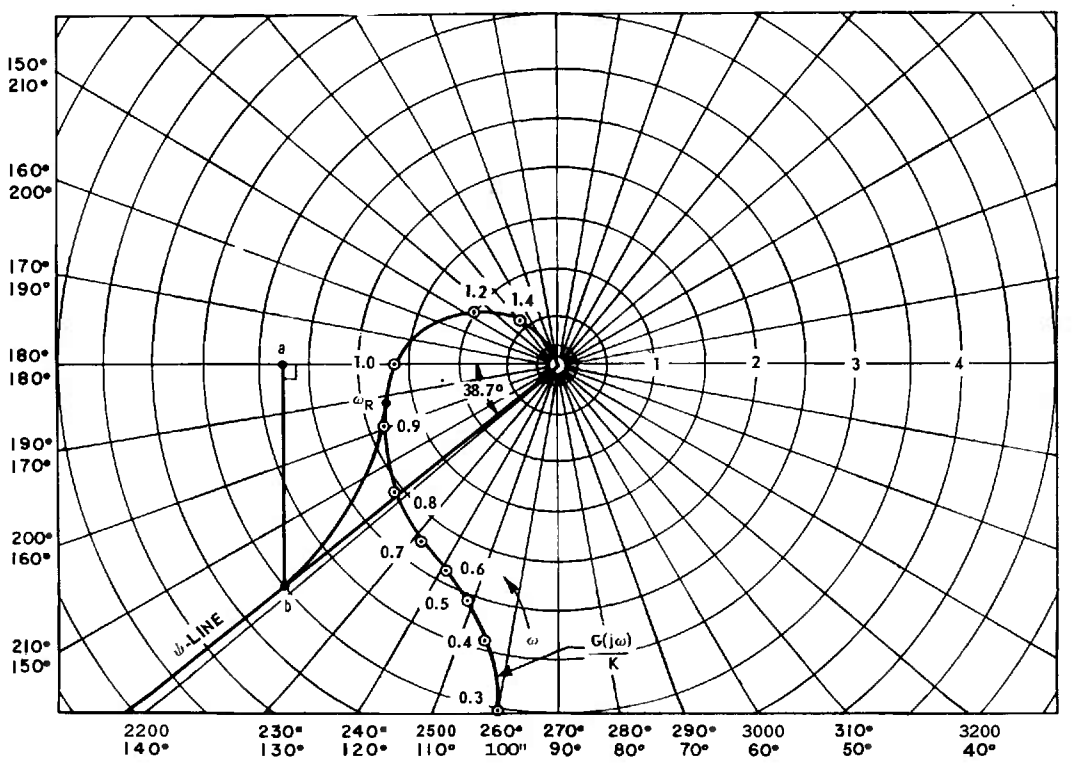


Fig. 5-25 Direct-plane determination of K for $M_p = 1.6$, $G(j\omega) = \dfrac{K}{j\omega[(j\omega)^2 + 0.6j\omega + 1]}$

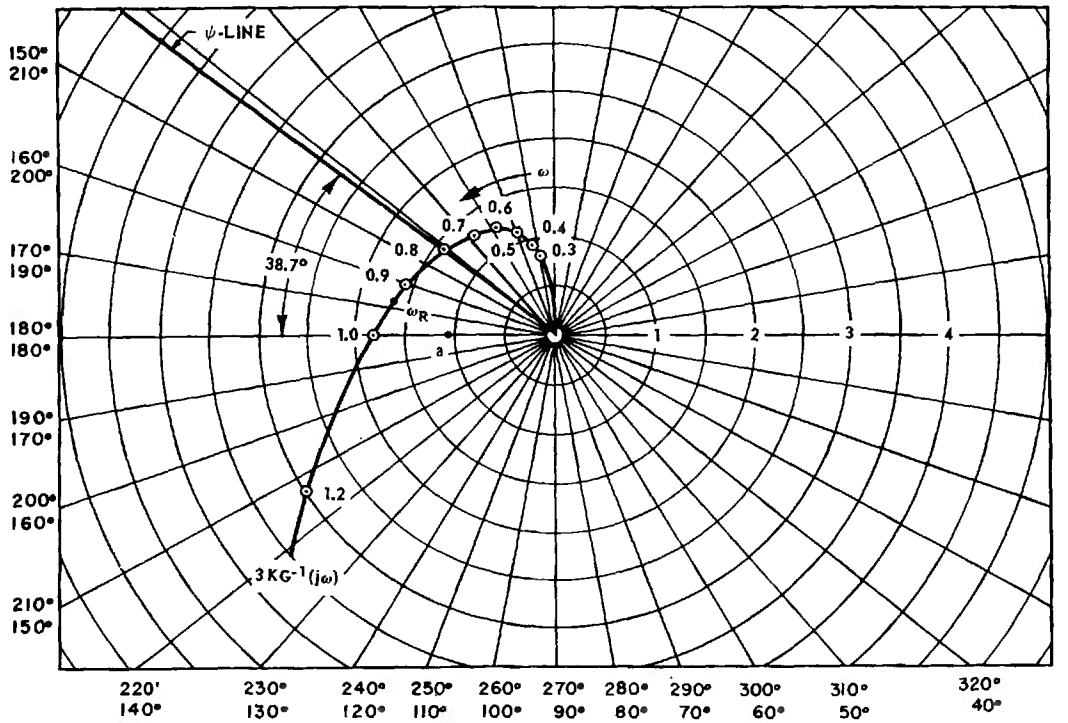


Fig. 5-26 Inverse-plane determination of K for $M_p = 1.6$, $G(j\omega) = \dfrac{K}{j\omega\,[(j\omega)^2 + 0.6j\omega + 1]}$.

(3) The circle with center on the real axis, tangent to the $G(j\omega)/K$ locus and the ψ line, is constructed.

(4) A line is drawn from point b perpendicular to the real axis, intersecting the real axis at the point $-2.78+j0$ (point a).

(5) Thus, $\dfrac{1}{K} = 2.78$, or $K = 0.36$.

(6) The resonant frequency is $\omega_R = 0.94$ rad/sec.

(b) Inverse-plane procedure :

(1) The locus of $3KG^{-1}(j\omega)$ is plotted (Fig. *5-26).*

(2) The ψ line for $M_p = 1.6$ ($\psi = 38.7°$) is drawn.

(3) The tangent circle is constructed.

(4) The center of the tangent circle is at $1.08+j0$ (point a).

(5) Thus, $3K = 1.08$, *or* $K = 0.36$.

(6) The resonant frequency is $\omega_R = 0.94$ rad/sec.

5-5.3 GAIN-PHASE PLANE CONSTRUCTION

The construction required to determine the gain K for a specified M_p is simpler when the gain-phase plane rather than the polar plane is employed. In the gain-phase plane construction, changing the gain merely moves the $G(j\omega)$ locus in a vertical direction without changing the phase angle. The gain-phase plane construction is carried out as follows (Fig. *5-27)* :

(a) The $G(j\omega)/K$ locus is drawn on the gain-phase plane.

(b) The $G(j\omega)/K$ locus is placed over the Nichols chart **or,** more specifically, **a** plot of the desired M_p contour is made on the gain-phase plane. The two coordinate systems are then aligned so that angles coincide.

(c) The $G(j\omega)/K$ locus is moved up **(or** down) until it is tangent to the specified M_p contour.

(d) The intersection of the 0-dg line of the M_p contour with the $G(j\omega)/K$ magnitude scale gives the value of $10 \log_{10} K$, where K is the value of the gain by which $G(j\omega)/K$ must be multiplied to produce the specified M_p.

(e) The angular frequency at the point of tangency of the M_p contour with the $G(j\omega)/K$ locus determines the resonant frequency of the closed-loop system having the specified M_p.

Example. A unity-feedback system has the open-loop frequency response

$$G(j\omega) = \frac{K\,(0.2j\omega + 1)}{j\omega\left[\left(j\frac{\omega}{10}\right)^2 + 0.6\,j\frac{\omega}{10} + 1\right]} \quad (5\text{-}36)$$

Find K and ω_R for $M_p = 1.5$.

Solution.

(a) The $\frac{G(j\omega)}{K}$ locus is plotted (Fig. 5-28).

(b) The $G(j\omega)/K$ locus is placed over the $M_p = 1.5$ contour, the phase-angle coordinates are aligned, and the locus is moved vertically until tangency occurs.

(c) The point of tangency occurs at $\omega_R = 12$ rad/sec.

(d) The intersection of the 0-dg line of the M_p contour with the magnitude scale of $G(j\omega)/K$ yields

$-10 \log_{10} K = -4.15$ dg, or $K = 2.6$

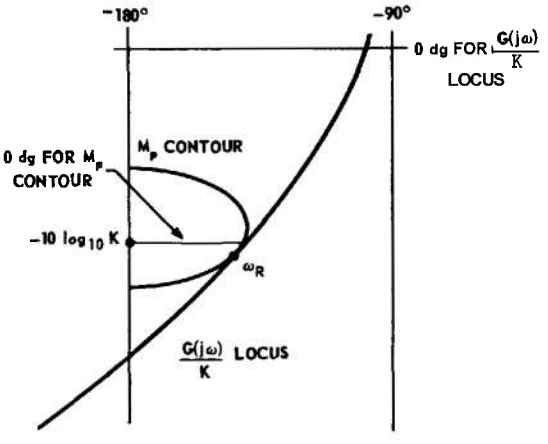


Fig. 5-27 Construction for gain determination on gain-phase plane.

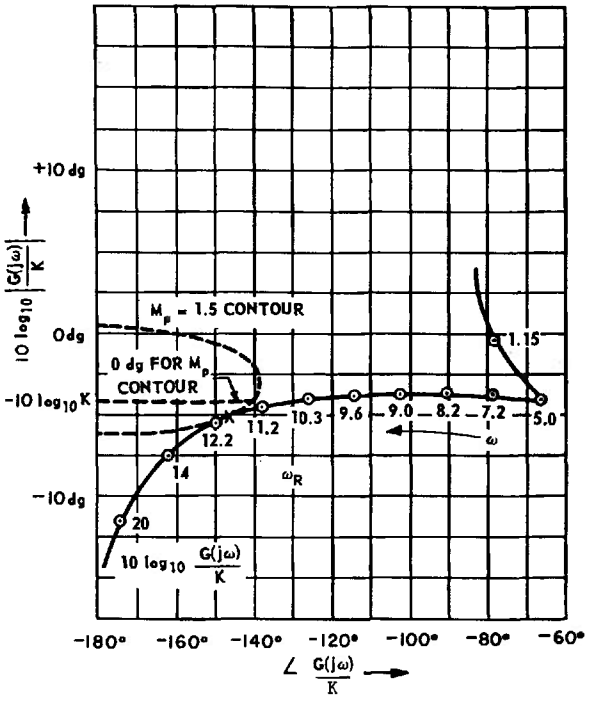


Fig. 5-28 Gain-phase plane determination of K for $M_p = 1.5$.

$$G(j\omega) = K\frac{(0.2\,j\omega + 1)}{j\omega\left[\left(j\frac{\omega}{10}\right)^2 + 0.6\,j\frac{\omega}{10} + 1\right]}$$

5-6 APPROXIMATE PROCEDURES[4,5,9]

5-6.1 PHASE MARGIN AND GAIN MARGIN

The peak magnitude of the closed-loop response is not the only measure of the degree of stability that is commonly used. More direct, but less reliable, descriptions of the approach of the $G(j\omega)$ locus to the point $-1+j0$ are available. These measures of the degree of stability are called *phase margin* and *gain margin*.

5-6.2 Phase Margin (Fig. 5-29)

The phase margin (p.m.) of the open-loop function $G(j\omega)$ of a unity-feedback system equals $[180° + \angle G(j\omega)]$ at the frequency for which the magnitude of $G(j\omega)$ is unity.

5-6.3 Gain Margin (Fig. 5-29)

The gain margin (g.m.) of the open-loop function $G(j\omega)$ of a unity-feedback system is the reciprocal of the magnitude of $G(j\omega)$ at the frequency for which the angle of $G(j\omega)$ is $-180°$.

The primary advantage of the use of the phase margin or the gain margin as a measure of the degree of stability is that calculations may be made directly on the separate magnitude and phase-angle plots.

In practice, the phase margin is used more widely than the gain margin as a degree-of-stability criterion, while the gain margin that results when the phase margin is specified is used as a measure of the goodness of performance. A system with a low gain margin is considered to have a poor performance.

The usual ranges of phase margin and gain margin for which performance will probably be satisfactory are the following:

$$30° < \text{p.m.} < 60° \qquad (5\text{-}37)$$

$$2.5 < \text{g.m.} < 10 \qquad (5\text{-}38)$$

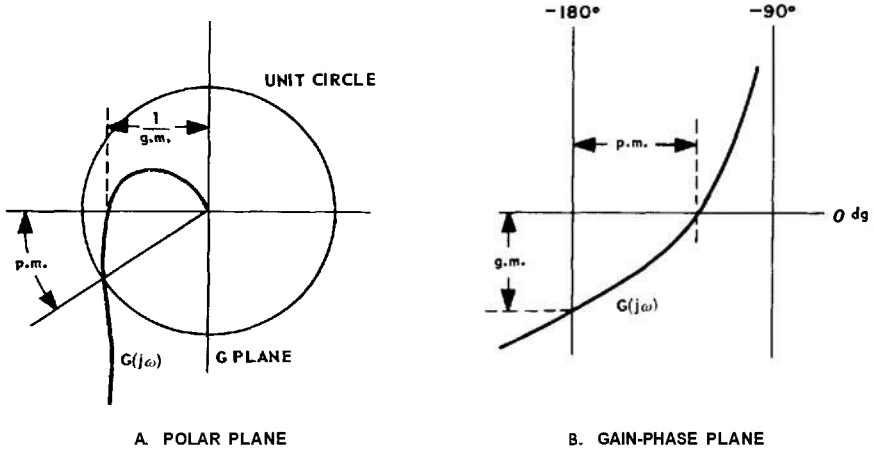
When the phase margin is used as a degree-of-stability criterion for setting the gain K of a unity-feedback system, the procedure is developed directly from the definition of phase margin as follows:

(a) The separate amplitude and phase plots (or the gain-phase plot) of $G(j\omega)/K$ are constructed.

(b) The frequency at which

$$\angle \frac{G(j\omega)}{K} = -180° + \text{p.m.} \qquad (5\text{-}39)$$

is determined.



A. POLAR PLANE B. GAIN-PHASE PLANE

Fig. 5-29 Phase margin and gain margin.

(c) At this frequency, $|G(j\omega)/K$ is determined.

(d) At this frequency, K is chosen such that

$$|G(j\omega)| = 1 \qquad (5\text{-}40)$$

Note that, if only a rough approximation is desired, the asymptotic magnitude curve may be used rather than the true magnitude curve.

Example.

The function plotted in Figs. 5-12 and 5-13 is the open-loop function of a unity-feedback system. The phase margin of the system is to be set at 45". In Fig. 5-13, $\omega = 11$ rad/sec when

$$\angle \frac{G(j\omega)}{K} = -180° + 45" = -135" \qquad (5\text{-}41)$$

In Fig. 5-11, for $\omega = 11$ rad/sec

$$10 \log_{10} \left| \frac{G(j\omega)}{K} \right| = -4.5 \text{ dg} \qquad (5\text{-}42)$$

To have $|G(j\omega)| = 1$ at $\omega = 11$ rad/sec

$$10 \log_{10} K = 4.5 \text{ dg, or } K = 2.82 \qquad (5\text{-}43)$$

Note, in this example, that the use of the asymptotic curve to estimate K gives a poor result. The magnitude of the asymptotic approximation for $G(j\omega)/K$ at $\omega = 11$ rad/sec is -7.5 dg. This would give an approximate value of $K = 5.61$ for a 45" phase margin. The error of approximation is a factor of two, which is too large to be acceptable. One should note further that, for this system, the gain margin is infinite since the negative phase shift never exceeds 180".

5-6.4 GENERAL COMMENTS ON THE PHASE-MARGIN CRITERION

The phase-margin criterion used as a measure of the degree of stability is a good substitute for the M_p criterion provided that the G function does not have low damping-ratio quadratic factors ($\zeta \leq 0.3$) with natural frequencies in the range where

$$-135" < \angle G(j\omega) < -225" \qquad (5\text{-}44)$$

If no low damping-ratio quadratics are present, then the gain determined from the true magnitude curve (or the asymptotic magnitude curve) for a specified phase margin is a good approximation to the gain determined from the corresponding M_p criterion. The M_p criterion corresponding to a given phase margin may be found from the relation

$$\text{p.m.} \approx \sin^{-1}\left(\frac{1}{M_p}\right) \qquad (5\text{-}45)$$

The frequency at which the phase margin is determined is: (1) the magnitude crossover frequency ω_{cm} if the true magnitude curve is used; and (2) the asymptote crossover frequency ω_c if the asymptotic magnitude curve is used. For the M_p corresponding to the specified phase margin, the frequencies ω_{cm} or ω_c are good approximations to the resonant frequency of the system ω_R.

5-6.5 APPROXIMATE CLOSED-LOOP RESPONSE

If the phase-margin criterion is used in conjunction with the separate amplitude and phase-angle plots, a rapid estimation of the closed-loop magnitude response is obtainable by means of the following relations:

(a) For a unity-feedback system (Fig. 5-1)

$$|W(j\omega)| \approx 1, \quad \text{when } |G(j\omega)| >> 1 \qquad (5\text{-}46)$$

$$|W(j\omega)| \approx |G(j\omega)|, \quad \text{when } |G(j\omega)| << 1 \qquad (5\text{-}47)$$

(b) For a nonunity-feedback system (Fig. 5-21)

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| \approx \left|\frac{1}{H(j\omega)}\right|, \quad \text{when } |G(j\omega)\,H(j\omega)| >> 1 \qquad (5\text{-}48)$$

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| \approx |G(j\omega)|, \quad \text{when } |G(j\omega)\,H(j\omega)| << 1 \qquad (5\text{-}49)$$

In the approximate equations (5-46) to (5-49), the boundary is always the point where the magnitude of the open-loop function is unity. Since this point is determined

directly in the phase-margin procedure, the approximate closed-loop response for a given phase-margin criterion can be constructed as follows:

(a) For unity feedback, the magnitude crossover frequency ω_{cm} is determined by means of the phase-margin criterion.

(b) Usually for $\omega < \omega_{cm}$, $|G(j\omega)| > 1$, and for $\omega > \omega_{cm}$, $|G(j\omega)| < 1$ [$G(j\omega)$ is monotonic].

(c) Therefore, for $\omega < \omega_{cm}$, $|W(j\omega)| \approx 1$. For $\omega > \omega_{cm}$, $|W(j\omega)| \approx |G(j\omega)|$.

(d) At $\omega = \omega_{cm}$, $|W(j\omega)| \approx M_p$, where M_p is determined for the specified phase margin from Eq. **(5-45).**

(e) From the high-frequency ($\omega > \omega_{cm}$) and low-frequency ($\omega < \omega_{cm}$) behavior together with the behavior at $\omega = \omega_{cm}$, the entire magnitude response $|W(j\omega)|$ may be approximated.

If $|G(j\omega)|$ is not monotonic as defined in step (b), several magnitude crossover points will exist and Eqs. **(5-46)** and **(5-47)** must be used directly. In this case, the approximation should not be trusted unless the crossover points are widely separated (at least 1 decade apart) or $|G(j\omega)| >> 1$ or $<< 1$ between the crossover points.

The procedure for nonunity-feedback systems is similar to that described here for unity-feedback systems and is based on Eqs. **(5-48)** and **(5-49).**

If only a very rough approximation is desired, the asymptotic magnitude curves may always be used to reduce calculation time.

Example. The open-loop function of a unity-feedback system is

$$G(j\omega) = \frac{K}{j\omega\,(j\omega + 1)} \qquad \textbf{(5-50)}$$

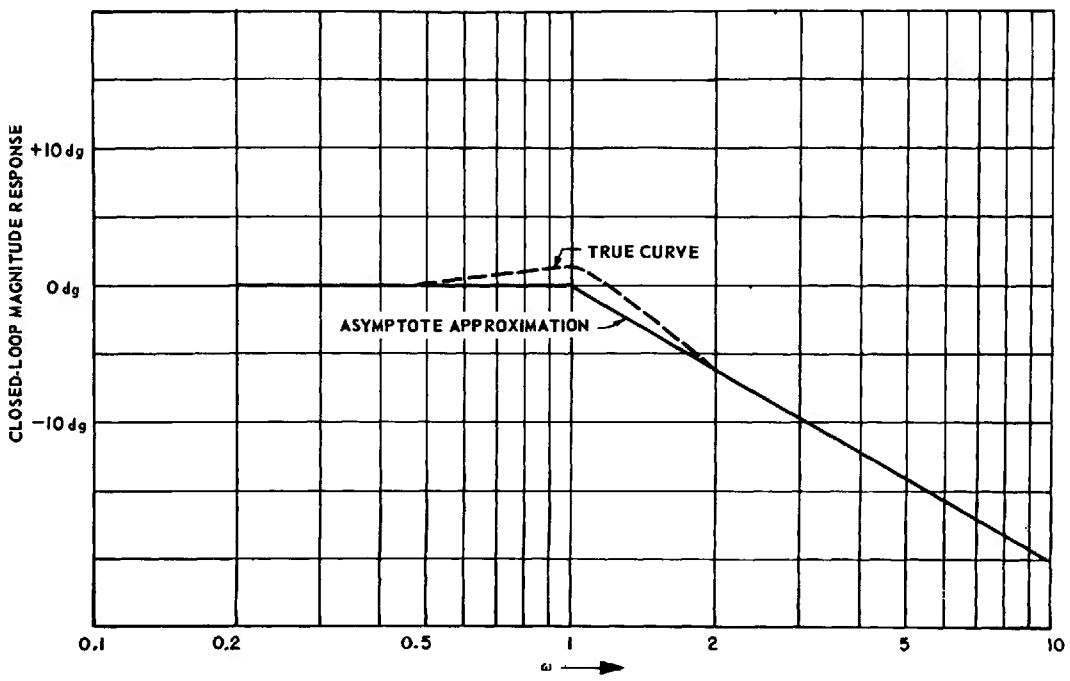


Fig. 5-30 Approximate closed-loop magnitude response of unity-feedback system,

$$G(j\omega) = \frac{K}{j\omega(j\omega + 1)}.$$

The gain K is to be set for a phase margin of 45". Find K, ω_{cm}, and the approximate closed-loop magnitude response $W(j\omega)$.

Solution.

(a) $\angle \frac{G(j\omega)}{K} = -180° + 45" = -135"$

for $\omega = 1$ (the crossover frequency).

(b) At $\omega = 1$, $\frac{G(j\omega)}{K} = 0.707$.

(c) Therefore, $K = 1.41$ and $\omega_{cm} = 1$.

(d) If the asymptotic magnitude curve for $G(j\omega)/K$ is used, the asymptotic magnitude of $G(j\omega)/K = 1$ at $\omega = 1$. Thus, use of the asymptotic magnitude rather than the true magnitude produces an error of 40 percent in the determination of K.

(e) If the approximation to the closed-loop magnitude response is based on the asymptotic magnitude curve of $G(j\omega)/K$, then for $\omega < \omega_{cm}$, the magnitude of $W(j\omega)$ is unity. For $\omega > \omega_{cm}$, the magnitude of $W(j\omega)$ is represented by a straight line with a slope of -20 dg/decade, crossing 0 dg at $\omega = \omega_{cm} = 1$. The M_p corresponding to a phase margin of 45° is 1.41, or 1.5 dg. The approximate closed-loop magnitude response is shown in Fig. 5-30.

5-7 ROOT-LOCUS METHOD[1-3]

5-7.1 GENERAL

The root-locus method deals primarily with the study of the motion in the s plane of the roots of the characteristic equation of a system as a function of the gain K. The relation between stability and gain can be observed directly through use of this method by noting how the roots move from the left half of the s plane (stable roots) to the right half of the s plane (unstable roots) as K is varied. A system can be characterized as having a low degree of stability if its roots lie in the left half of the s plane but are very close to the imaginary axis.

Gain determination by means of the root-locus method is based on the fact that many practical systems have a pair of complex closed-loop poles that are closer to the origin than any other complex poles of the system. These poles are called the ***dominant*** poles of the system. By assigning a specified value to a characteristic parameter of the dominant pole pair, the gain of the system may be fixed by a measure of the degree of stability related to the dominant pole pair.

5-7.2 PROPERTIES OF ROOTS IN THE s PLANE

The roots of the characteristic equation of a system are either first- or second-order, and each root is associated with a specific transient response mode. The characteristics of the roots, the response modes, and the specific contours in the s plane are related in a simple way.

5-7.3 First-Order Root: $s = -\frac{1}{T}$

The transient response mode corresponding to this root is $e^{-t/T}$, where T is the time constant of the mode. Lines drawn parallel to the imaginary axis in the s plane are loci of constant T for first-order factors.

5-7.4 Second-Order Root:

$$s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$$

The transient response mode corresponding to this root is

$$e^{-\zeta\omega_n t}\cos\omega_d t$$

where

$(\zeta\omega_n)^{-1}$ = time constant of envelope of mode

$\omega_d = \omega_n\sqrt{1-\zeta^2}$ = damped frequency of transient oscillation

ζ = damping ratio

ω_n = undamped natural frequency

For the second-order root, s-plane loci can be developed by using the following properties:

$$Re(s) = -\zeta\omega_n \quad (5\text{-}51)$$

$$Im(s) = \pm\omega_d = \pm\omega_n\sqrt{1-\zeta^2} \quad (5\text{-}52)$$

$$|s| = \omega_n \quad (5\text{-}53)$$

$$\angle s = \pm\cos^{-1}\zeta \pm 180" \quad (5\text{-}54)$$

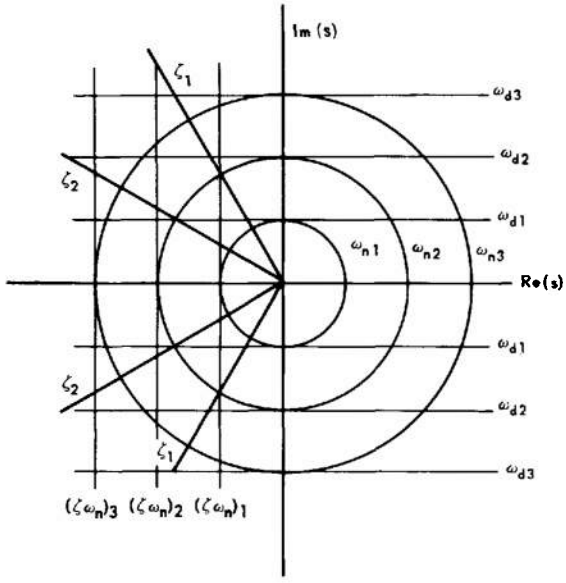


Fig. 5-37 loci of characteristic parameters of second-order root.

Thus, lines drawn parallel to the imaginary axis are loci of constant-envelope time constant $(\zeta\omega_n)^{-1}$. Lines drawn parallel to the real axis are loci of constant damped frequency of oscillation ω_d. Circles centered at the origin are loci of constant natural frequency ω_n. Radial lines emanating from the origin are loci of constant damping ratio ζ. The various s-plane loci for the second-order factor are shown in Fig. 5-31.

5-7.5 GAIN DETERMINATION IN THE s PLANE

The usual degree-of-stability criterion for determining the gain K from the root locus of a system is:

The dominant roots are adjusted to satisfy a specified damping ratio ζ.

The advantage of the root-locus procedure over the M_p criterion of the frequency-response method becomes evident if a truly dominant pole pair exists. In this case, all other poles are far from the origin, and the transient response of the system is dominated by the transient response mode associated with the dominant pole pair. Thus, the time behavior of the system becomes immediately evident once the damping-ratio criterion is satisfied. The disadvantage of the root-locus method is that considerable time is consumed in constructing the locus. Adjustment of the gain K for a specified dominant-root damping ratio is best demonstrated by a specific example.

Example. The open-loop function $G(s)$ of a unity-feedback system is

$$G(s) = \frac{K(0.2s + 1)}{s(s + 1)(0.1s + 1)} \qquad (5\text{-}55)$$

Find the gain K and the closed-loop pole-zero configuration for a dominant-root damping ratio $\zeta = 0.5$.

Solution

(a) The open-loop function is placed in the standard form of the root-locus method as follows:

$$G(s) = 2K \frac{(s + 5)}{s(s + 1)(s + 10)} \qquad (5\text{-}56)$$

(b) The angle condition is

$$\angle(s + 5) - \angle s - \angle(s + 1) - \angle(s + 10) = -180° \qquad (5\text{-}57)$$

(c) The magnitude condition is

$$2K \frac{|s + 5|}{|s|\,|s + 1|\,|s + 10|} = 1 \qquad (5\text{-}58)$$

(d) The locus of the roots is constructed from the angle condition by choosing trial points in the s plane and checking back to see whether the angles of the vectors of the open-loop poles and zeros add up according to Eq. (5-57). A curve drawn through the trial points that satisfy this equation is the root locus. To determine the gain K associated with each locus point, Eq. (5-58) is used. The value of s corresponding to a given locus point is substituted into this magnitude equation and the equation is then solved for K. The locus for this problem in the upper half of the s plane is shown in Fig. 5-32.

(e) The line corresponding to $\zeta = 0.5$ ($\angle s = \pm 120°$) is drawn and the intersection with the locus is noted. The intersection occurs at $s = -0.60 \pm j1.04$.

(f) Using the magnitude condition, it is found that $K = 1.38$.

(g) For $K = 1.38$, the location of the real root between $s = -5$ and $s = -10$ is determined by direct application of the magnitude condition. This root lies at $s = -9.85$.

(h) The closed-loop pole-zero configuration has a zero at the open-loop zero $s = -5$.

(i) The factored closed-loop transfer function is

$$W(s) = \frac{2.84\ (s + 5)}{(s + 9.85)\ (s^2 + 1.2s + 1.44)}$$

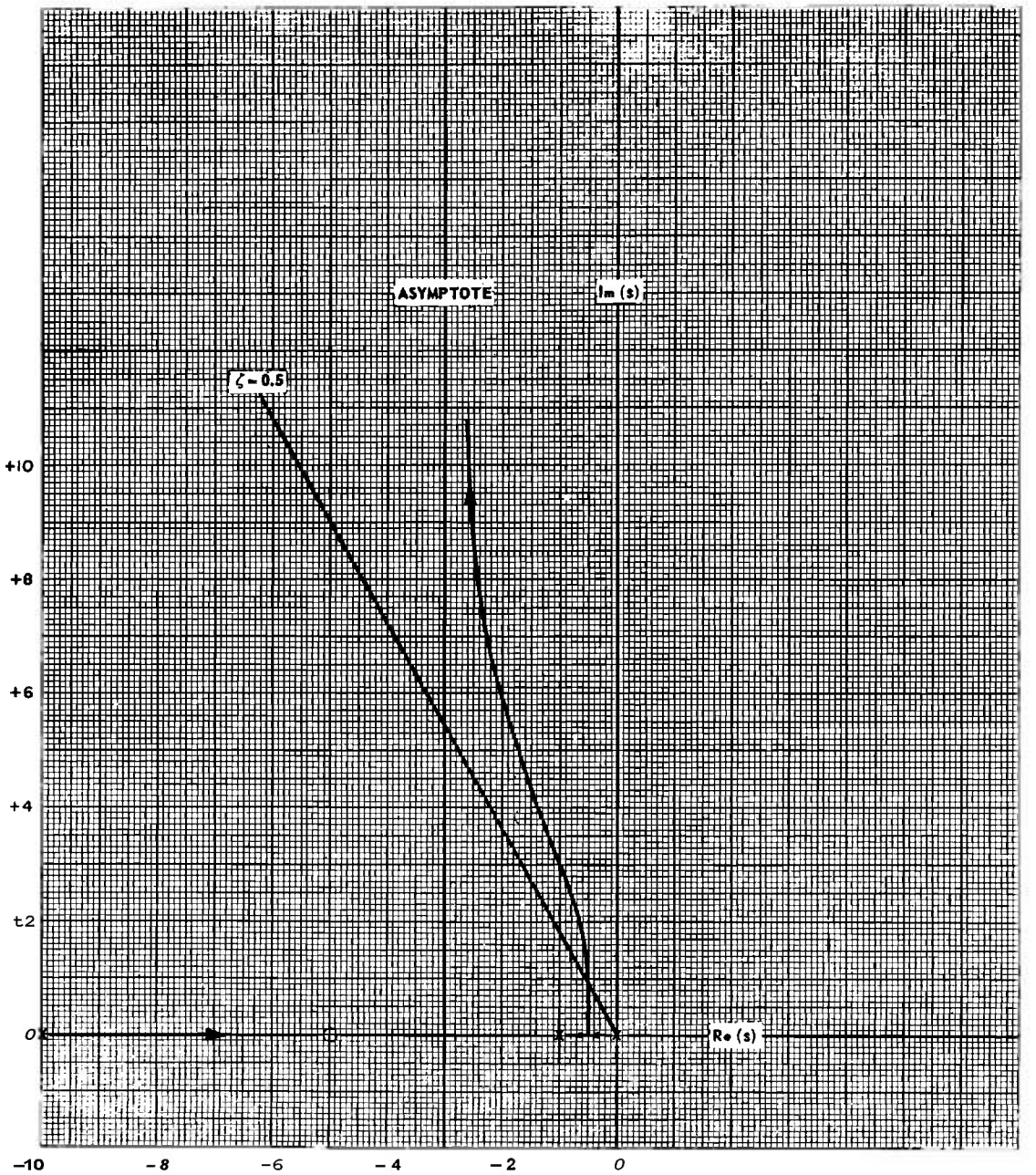


Fig. 5-32 Gain determination from root locus, $G(s) = \dfrac{K(0.2s + 1)}{s(s + 1)\ (0.1s + 1)}$.

BIBLIOGRAPHY

1 G. S. Brown and D. P. Campbell, ***Principles of Servomechanisms,*** pp. #92-104, John Wiley & Sons, Inc., New York, N.Y., 1948.

2 ibid., pp. #146-154.

3 ibid., pp. #166-168.

4 ibid., pp. #176-189.

5 ibid., pp. #236-261.

6 J. G. Truxal, ***Automatic Feedback Control System Synthesis,*** pp. #80-81, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

7 ibid., pp. #221-242.

8 W. R. Evans, ***Control-System Dynamics,*** pp. #96-113, McGraw-Hill Book Company, Inc., New York, N.Y., 1954.

9 H. M. James, N. B. Nichols, and R. S. Phillips, ***Theory of Servomechanisms,*** MIT Radiation Laboratory Series, Vol. 25, pp. #158-186, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

10 W. R. Ahrendt and J. F. Taplin, ***Automatic Feedback Control,*** McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

11 Edited by I. A. Greenwood, Jr., J. V. Holdam, Jr., and D. MacRae, Jr., ***Electronic Instruments,*** MIT Radiation Laboratory Series, Vol. 21, pp. #215-318, McGraw-Hill Book Company, Inc., New York, N. Y., 1948.

12 H. Chestnut and R. W. Mayer, ***Servomechanisms and Regulating System Design,*** Vol. I, pp. #221-244, John Wiley & Sons, Inc., New York, N. Y., 1951.

13 ibid., pp. #291-326.

14 ibid., pp. #347-350.

CHAPTER 6

COMPENSATION TECHNIQUES *

6-1 INTRODUCTION

Compensation in the general field of servomechanisms refers to the procedures used to modify the dynamic response characteristics of a system by auxiliary means so that it meets performance specifications. Most actual components have a limited dynamic response and so do not respond instantaneously to input variations. Because of the dynamic limitations of physical components, stability problems arise in closed-loop systems, as discussed in Chs. **4** and **5.** The requirement of stable operation imposed on all closed-loop systems limits the accuracy that can be obtained with these systems. In order to minimize this inherent limitation, artificial means (compensation techniques) are used to modify the dynamic characteristics. These include the introduction of networks cascaded with the fixed elements in the loop and the addition of auxiliary loops to the system.

The general compensation problem is illustrated by Fig. 6-1. In this figure, $G_f(s)$ represents the response of the fixed elements in the loop which cannot be altered, $H(s)$ represents the response of feedback elements that may be present, and $G_c(s)$ represents the response of compensating elements that are to be adjusted so that the complete system meets the performance specifications. The procedure for designing the system can be outlined as follows [H(s) is assumed to be unity] :

(a) With $G_c(s) = K$, a pure gain (real number), a stability check is made to determine the allowable range of the gain K for stable operation.

(b) Assuming a specified degree of stability, the gain K, is adjusted to meet this requirement.

(c) From the input specifications, the error of the system is found when K is adjusted as in (b), and this error is checked against the error specification.

(d) If the error does not meet specifications, a more complicated form for $G_c(s)$ is introduced. The system is then adjusted to satisfy the specified 'degree of stability, and the error specification is again checked.

(e) The procedure is continued, trying different or more complicated compensation functions, until the error falls within specifications (that is, if the specifications can be met).

In practice, the forms of compensation normally employed are kept simple. This is due in part to the fact that the fixed elements are usually limited in their range of linear operation, and the introduction of complex compensation functions merely reduces the range over which the linearity assumption applies. In addition, it is found that the theoretical advantages that may accrue with complex compensation are not realized in practice because the theoretical model no longer fits the physical system.

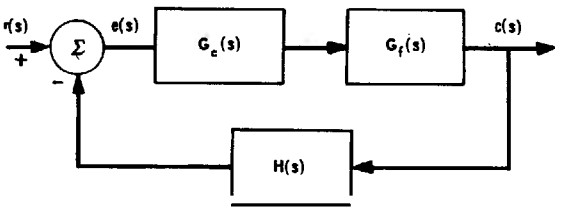


Fig. 6-7 Compensation in a single loop.

By L.A. Gould

6-2 RESHAPING LOCUS ON GAIN-PHASE PLANE [1,2,3,4,8,11]

6-2.1 GENERAL

Because magnitudes and phase angles add when functions are cascaded in the gain-phase plane, this frequency domain is the one most suitable for studying the effects of cascaded compensation functions. The introduction of a compensation function in the loop of a unity-feedback system can be thought of as a method for reshaping the open-loop frequency response $C(j\omega)/E(j\omega)$ to permit a higher gain to be used for the specified degree of stability. If M_s is the degree-of-stability criterion used, the gain can be increased by causing the phase and gain margins of the function $G_c(j\omega)G_f(j\omega)/K$ to increase by a proper choice of $G_c(j\omega)$. To maintain the specified M_s, the M_p contour must be moved *downward* for tangency to occur; this downward motion corresponds to an increase in the open-loop gain of the system (Fig. 6-2). The two most commonly used compensation networks for reshaping the open-loop gain-phase locus are the lag network and the lead network (see Par. 6-6).

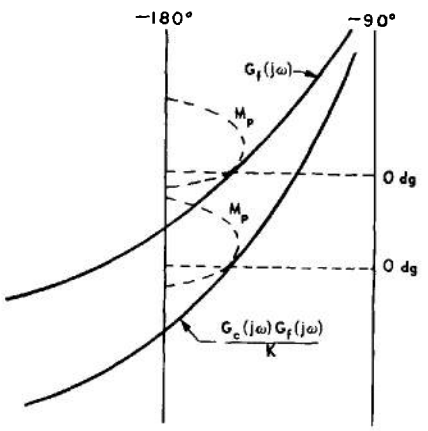


Fig. 6-2 Change in open-loop response produced by compensation illustrating downward motion of M_p contour for gain increase

6-2.2 LAG COMPENSATION

The first-order lag function is

$$G_c(j\omega) = K\,\frac{T_c j\omega + 1}{\alpha T_c j\omega + 1}\;; \quad a > 1 \qquad (6\text{-}1)$$

where K is the real gain factor, T_c is the time constant, and a is an attenuation factor. Lag-function plots for $a = 5, 10, 20$ (with $K = 1$) are shown in Fig. 6-3. These plots are made to a normalized frequency scale for which

$$\beta = T_c\omega \qquad (6\text{-}2)$$

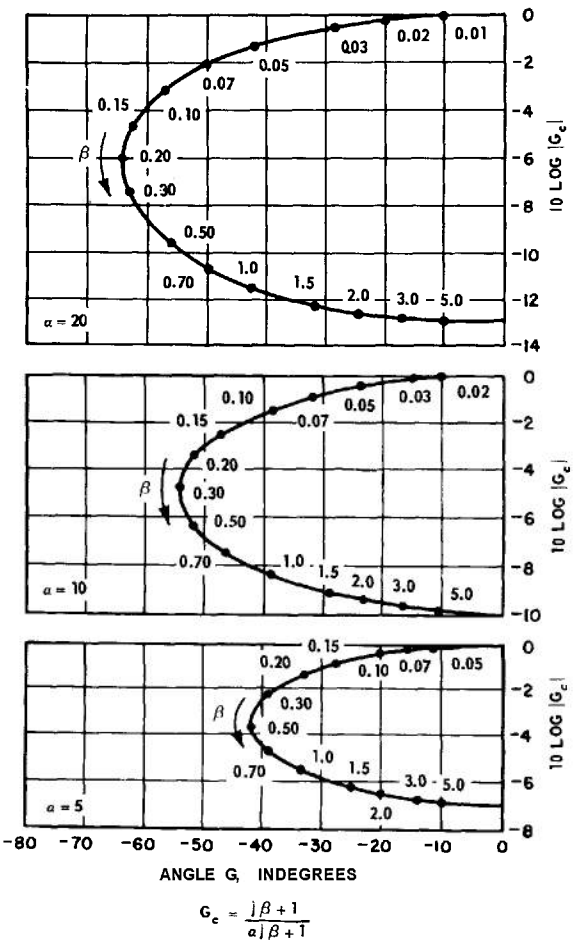


Fig. 6-3 Universal lag functions.

In using the lag function to reshape an open-loop frequency response, the choice of the attenuation factor is usually governed by the gain increase that is sought. In practice, a's greater than **20** are rarely used. **A** good rule of thumb is that the gain increase that can be achieved lies in the range from 0.7α to 0.9α. There are two considerations that limit the choice of the time constant T_c. Since the lag function introduces a negative phase shift, the choice of the time constant should be such that this phase shift does not occur in the region where the open-loop response passes near the $-1+j0$ point. Consequently, the lag function is usually adjusted so that its major phase contribution occurs at low frequencies. This means that T_c cannot be made too small without affecting the stability of the system. On the other hand, if T_c is made too large, the transient response of the system tends to deteriorate as a result of excessive peaking and an abnormally long settling time. **A** rule of thumb for adjusting the lag function is as follows: Choose T_c so that a phase shift of **—5°** to $-10°$ is introduced at the uncompensated resonant frequency of the system. The uncompensated resonant frequency is defined as the resonant frequency which is obtained for the specified M_p when $G_c(s) = K$.

The steps involved in adjusting the lag function are the following:

(a) With $G_c(s) = K$, the gain and resonant frequency of the system are found for the specified value of M_p.

(b) Using Fig. **6-3,** the value of β is determined for the specified allowable phase shift. Since the region of low phase shift of the lag function occurs for values of $\beta \geqq 5$, the phase shift of the lag function in this region is adequately represented by

$$\angle G_c(j\beta) \cong -\left(\frac{\alpha-1}{\alpha}\right)\frac{1}{\beta} \text{ for } \beta \geqq 5 \quad \dagger \qquad (6\text{-}3)$$

† Symbol $\angle$ denotes "angle"

Thus, the phase angle varies inversely as β.

(c) If ω_{R1} is the uncompensated resonant frequency and β_ϕ is the value of β corresponding to the specified allowable phase shift of the lag function, then, from Eq. **(6-2),**

$$T_c = \frac{\beta_\phi}{\omega_{R_1}} \qquad (6\text{-}4)$$

(d) Since the scale ratio between β and ω is fixed by Eq. **(6-4),** the magnitude and phase-angle contribution of the lag function to the $G_f(j\omega)$ function at each frequency can determined from the universal curves of Fig. **6-3.**

(e) After the lag function has been added to the fixed-element response $G_f(j\omega)$, the gain K is determined from the specified M_p criterion.

Example. The transfer function of the fixed elements of a unity-feedback system is given by

$$G_f(s) = \frac{1}{s(0.3s+1)(0.1s+1)} \qquad (6\text{-}5)$$

A lag function with $a = 10$ is used to compensate the system. The lag function is to contribute **—5°** of phase shift at the uncompensated resonant frequency. Design the compensation when $10\log_{10} M_p = 1.5$ dg.

Solution. The frequency response $G_f(j\omega)$ is plotted in Fig. **6-4** as Curve A. For the specified M_p (see construction), the point of tangency of the M_p contour with Curve A occurs at the point where $\omega_{R_1} = 2.4$ rad/sec; from the displacement downward of the M_p contour by **4.1** dg (i.e., **10** log $K = 4.1$) one gets $K = 2.57$ for the uncompensated system when $G_c(j\omega) = K$. From the $a = 10$-plot of Fig. **6-3,** —5° of phase shift occurs at $\beta_\phi = 10$. Therefore,

$$T_c = \frac{\beta_\phi}{\omega_{R_1}} = \frac{10}{2.4} = 4.17 \text{ seconds} \qquad (6\text{-}6)$$

The scale change from ω to β is, therefore,

$$\beta = 4.17\omega \qquad (6\text{-}7)$$

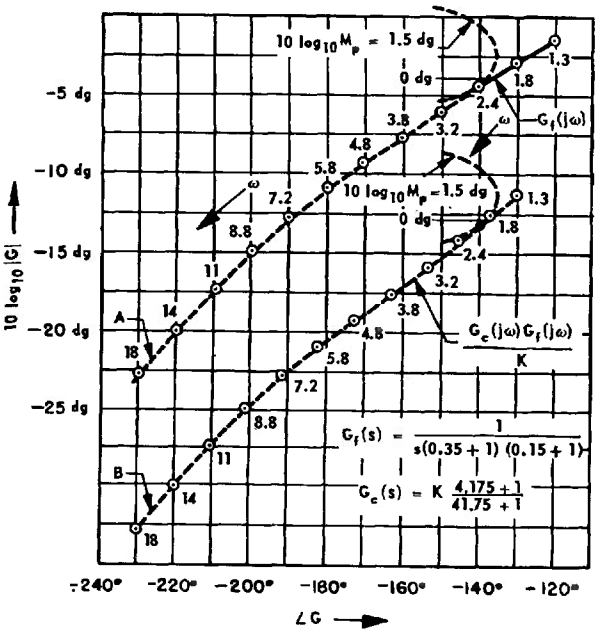


Fig. 6-4 lag-compensation procedure.

With this scale change, the universal lag function for $a = 10$ is used to add the magnitude and phase angle of $G_c(j\omega)/K$ to the fixed-element response $G_f(j\omega)$. The composite $G_c(j\omega)G_f(j\omega)/K$ function appears as Curve B in Fig. 6-4. Using the specified M_p criterion, the resonant frequency of the compensated system is 2.0 rad/sec and K has been increased to a value $K = 20.4$. Thus, the use of the lag function with $a = 10$ has allowed an increase in gain by a factor of 7.6. The resonant frequency has been decreased by 17 percent.

6-2.3 LEAD COMPENSATION

The first-order lead function is

$$G_c(j\omega) = K\,\frac{\alpha T_c j\omega + 1}{T_c j\omega + 1}\;; \qquad \alpha > 1 \qquad (6\text{-}8)$$

where K is the gain, T_c is the time constant, and a is an attenuation factor. Lead-function plots for $a = 5$, 10, and 20 are shown in Fig. 6-5. These plots utilize a normalized frequency scale for which

$$\beta = T_c\omega \qquad (6\text{-}9)$$

In using the lead function to reshape an open-loop frequency response, advantage is taken of the fact that the lead function exhibits positive phase shift. Thus, by adjusting the time constant T_c, it is possible to add positive phase angles to the fixed-element response $G_f(j\omega)$ in a region where the negative phase shift of the fixed elements is too great to secure an M_p-contour tangency. Hence, the lead func-

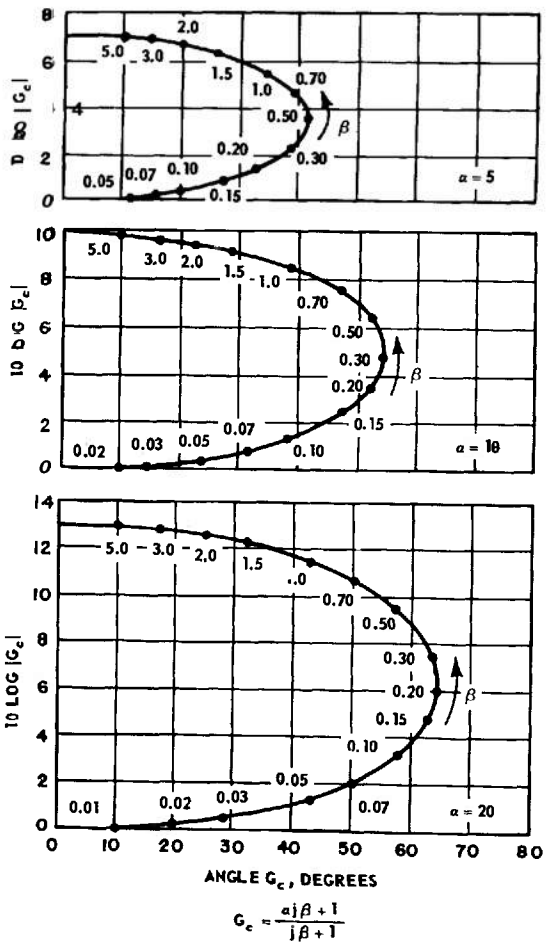


Fig. 6-5 Universal lead functions.

tion can decrease the effective negative phase shift of the composite $G_c(j\omega)G_f(j\omega)/K$ function, allowing an M_p tangency to occur for a larger value of gain K. In many cases, an increase in the resonant frequency of the system is obtained when lead compensation is used.

Due to the tendency of lead compensation to increase the bandwidth of a system, it is found that the system is more sensitive to noise, and its linear range of operation is restricted. Thus, in practical situations, the attenuation factor a used usually does not exceed **20.** The adjustment of the time constant T_c is a matter of trial and error. An outline of the trial-and-error procedure is given below :

(a) The phase-angle difference ψ between the "nose" of the M_p contour and the **−180°** line of the gain-phase plane is given by

$$\psi = \sin^{-1}\left(\frac{1}{M_p}\right) \qquad (6\text{-}10)$$

The maximum phase shift ϕ_m introduced by the lead function is

$$\phi_m = \sin^{-1}\left(\frac{a-1}{\alpha+1}\right) \qquad (6\text{-}11)$$

(b) The first-trial choice of the lead-function time constant T_c is found by determining the frequency at which the following angle relation holds :

$$\angle G_f(j\omega_1) = -180° + \psi - \phi_m \qquad (6\text{-}12)$$

The frequency ω_1 which satisfies Eq. **(6-12)** be found directly from the gain-phase plot of $G_f(j\omega)$. Then, the first choice of T_c is given by

$$T_{c_1} = \frac{\beta_m}{\omega_1} \qquad (6\text{-}13)$$

where β_m is the frequency at which the maximum phase shift (ϕ_m) occurs on the normalized lead-function curve of **Fig. 6-5.** The frequency β_m can be found from the relation

$$\beta_m = \frac{1}{\sqrt{\alpha}} \qquad (6\text{-}14)$$

(c) Since the scale ratio between ω and the normalized frequency β is fixed by Eq. **(6-14),** the magnitude and phase-angle contribution **of** the lead function to the $G_f(j\omega)$ function at each frequency can be determined from the universal curves of Fig. **6-5.**

(d) After the lead function has been added to the fixed-element response $G_f(j\omega)$, the gain K is determined from the specified M_p criterion.

(e) The first-trial choice of the lead-function time constant T_c usually determines a closed-loop resonant frequency that is close to the maximum obtainable frequency with the given attenuation factor a.

However, the open-loop gain K is not necessarily a maximum. Therefore, if gain increase is the objective, additional trials must be made. The additional trials usually involve the choice of trial values of T_c that are smaller than the initial choice. **A** rule of thumb is that the time constant T_c which maximizes K is approximately one-half to one-third the initial-trial value.

Example. The transfer function of the fixed elements of a unity-feedback system is given by Eq. **(6-5).** A lead function with $a = 10$ is used to compensate the system. Design the compensation when $X \log_{10} M_p = 1.5$dg.

Solution. The frequency response $G_f(j\omega)$ is plotted in Fig. **6-6** (Curve A) . From the problem specifications and Eqs. **(6-10), (6-11),**

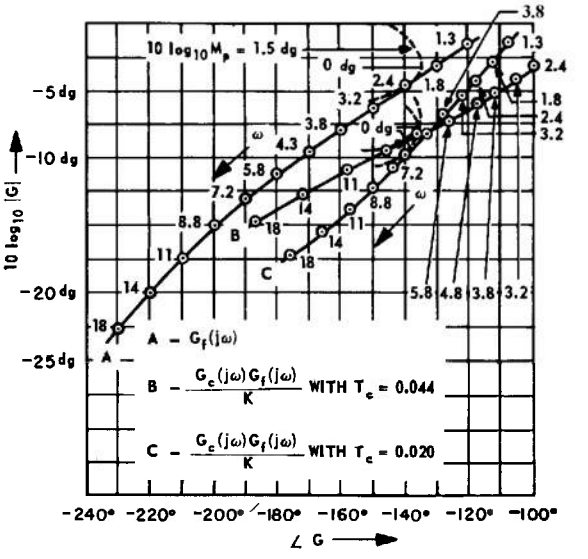


Fig. 6-6 Lead-compensation procedure.

and (6-14), we find that $\psi = 45°$, $\phi_m = 54.9°$, and $\beta_m = 0.316$. Eq. (6-12) yields

$$\angle G_f(j\omega_1) = -189.9° \qquad (6\text{-}15)$$

Using this result and Curve *A* of Fig. 6-6, we find that $\omega_1 = 7.2$ rad/sec. Then, from Eq. (6-13), the initial choice for the lead-function time constant is $T_{c_1} = 0.044$ sec. The scale change from ω to β is therefore given by

$$\beta = 0.044\omega \qquad (6\text{-}16)$$

With this scale change, the universal lead function for $a = 10$ is used to add the magnitude and phase angle of $G_c(j\omega)/K$ to the fixed-element response $G_f(j\omega)$. The composite $G_c(j\omega)G_f(j\omega)/K$ function appears as Curve *B* in Fig. 6-6. Using the specified M_p criterion, the resonant frequency of the uncompensated system (Curve *A*) is $\omega_{R_1} = 2.4$ rad/sec and the gain of the uncompensated system with $G_c(j\omega) = K$ is $K = 2.57$. The resonant frequency of the compensated system (Curve *B*) is $\omega_{R_2} = 8.2$ rad/sec and the gain of the compensated system is $K = 6.68$. Thus, the resonant frequency of the system has been increased by a factor of 3.4 and the gain by a factor of 2.6 through lead compensation. If maximum gain is sought, the lead time constant must be reduced. By trial and error, we find that with $T_c = 0.020$ sec, $\omega_R = 5.6$ rad/sec and $K = 8.65$. The construction for maximum gain is shown as Curve *C* in Fig. 6-6. Thus, for maximum gain adjustment, lead compensation increases the resonant frequency by a factor of 2.3 and the gain by a factor of 3.4. Note that the time constant for maximum gain is 0.45 times the initial-trial choice. The results of this example are listed in Table 6-1.

Compensation Adjustment	T_c (sec)	ω_R (rad/sec)	K
Initial trial ; maximum resonant frequency	0.044	8.2	6.68
Final trial ; maximum gain	0.020	5.6	8.65
No compensation		2.4	2.57

6-3 PHASE-MARGIN AND ASYMPTOTIC METHODS [3,4,8]

6-3.1 GENERAL

A rough picture of the effect of compensation can be obtained if the magnitude asymptotes are used in conjunction with the phase-margin criterion for the degree of stability (see Par. 5-6). Using a 45° phase-margin criterion, the asymptotic method gives good results provided there are no low-damping-ratio quadratic factors in the open-loop transfer function $C(j\omega)/E(j\omega)$. If the 45° phase-margin criterion is assumed, then the asymptote crossover frequency ω_c (defined in Par. 5-1) usually occurs in a region where the slope of the asymptote is −10 dg/decade.

The phase-margin criterion can be used as an approximation to the M_p criterion, or it can be used as an independent degree-of-stability criterion. If it is used independently, the separate magnitude vs frequency and phase-angle vs frequency plots are employed. If the phase-margin criterion is used to approximate the M_p criterion, one might just as well use the asymptotic curve as an approximation to the true magnitude curve.

The rules of thumb established in Par. 6-2 for adjusting the compensation functions can be applied in an identical manner when working with the separate magnitude and phase angle plots except that the asymptotic-crossover frequency ω_c or the magnitude-crossover frequency ω_{cm} will replace the resonant frequency ω_R where necessary. The universal compensation-function curves of Figs. 6-3 and 6-5 can be used when working with the separate magnitude and phase-angle curves.

The steps involved in adjusting the compensation functions when using the phase-margin criterion and the separate response curves are outlined below.

6-3.2 LAG COMPENSATION

(a) With $G_c(s) = K$, the gain and magnitude-crossover frequency (or asymptote-crossover frequency) are found for the specified phase margin.

(b) Using the universal lag-function curve (Fig. 6-3), the normalized frequency β_ϕ (at which the allowable negative phase shift of the lag function occurs) is determined.

(c) The lag-function time constant is found from the relation

$$T_c = \frac{\beta_\phi}{\omega_{cm_1}} \qquad (6\text{-}17)$$

where ω_{cm_1} is the magnitude-crossover frequency of the uncompensated system, or, alternatively, one can use the relation

$$T_c = \frac{\beta_\phi}{\omega_{c_1}} \qquad (6\text{-}18)$$

where ω_{c_1} is the asymptote-crossover frequency of the uncompensated system.

(c) Using the scale ratio between β and ω determined by Eq. (6-17) or (6-18), the magnitude (or asymptote) and phase-angle contribution of the lag function to the $G_f(j\omega)$ function at each frequency can be determined.

(d) After the lag function has been added to the fixed-element response $G_f(j\omega)$, the gain K is determined from the specified phase margin.

6-3.3 LEAD COMPENSATION

(a) The frequency ω_1 which satisfies Eq. (6-13) is found directly from the $\angle G_f(j\omega)$ curve.

(b) If β_m is the normalized frequency at which the maximum phase shift ϕ_m for the lead function occurs (see Fig. 6-5), then the first trial choice of the lead-compensation time constant is

$$T_{c_1} = \frac{\beta_m}{\omega_1} \qquad (6\text{-}19)$$

(c) Using the scale ratio between β_m and ω_1 determined by Eq. (6-19), the magnitude (or asymptote) and phase-angle contribution of the lead function to the $G_f(j\omega)$ function at each frequency can be determined.

(d) After the lead function has been added to the fixed-element response $G_f(j\omega)$, the gain K is determined from the specified phase margin.

(e) Further trial values of the lead-function time constant T_c are tried until the gain K or the magnitude- (or asymptote-) crossover frequency has been maximized.

Example. The transfer function of the fixed elements of a unity-feedback system is given by Eq. (6-5). The 45° phase-margin criterion is to be used to adjust the degree of stability of the system.

(a) Lag compensation with $a = 10$ is to be used to improve performance. The allowable negative phase shift that the lag function contributes at the magnitude- (or asymptote-) crossover frequency of the uncompensated system shall be 5". Design the compensation.

(b) Lead compensation with $a = 10$ is to be used to improve performance. Design the compensation.

Solution.

(a) The magnitude and asymptote curves of $G_f(j\omega)$ are drawn in Fig. 6-7 and the phase angle curve is drawn in Fig. 6-8. For a 45" phase margin, the magnitude-crossover frequency $\omega_{cm} = 2.08$ rad/sec and the corresponding gain is 2.34. The asymptote-crossover frequency $\omega_c = 2.08$ rad/sec and the corresponding gain is 2.04. (Compare with $\omega_R = 2.4$ rad/sec and $K = 2.57$ for 10 $\log_{10} M_p = 1.5$ dg.) For 5° of allowable negative

phase shift, and $a = 10$, $\beta_\phi = 10$ from Fig. 6-3. Therefore, using Eqs. (6-17) or (6-18), $T_c = 4.8$ sec. The scale change from β to ω is given by

$$\beta = 4.8\omega \qquad (6\text{-}20)$$

The composite $G_c(j\omega)G_f(j\omega)/K$ magnitude and asymptote curves for the lag-compensated system are drawn in Fig. 6-7, and the composite phase-angle curve is drawn in Fig. 6-8. Using the 45° phase margin, the magnitude-crossover frequency $\omega_{cm} = 1.74$ rad/sec, and the corresponding gain is 20. The asymptote-crossover frequency $\omega_c = 1.74$ rad/sec and the corresponding gain is 17.8.

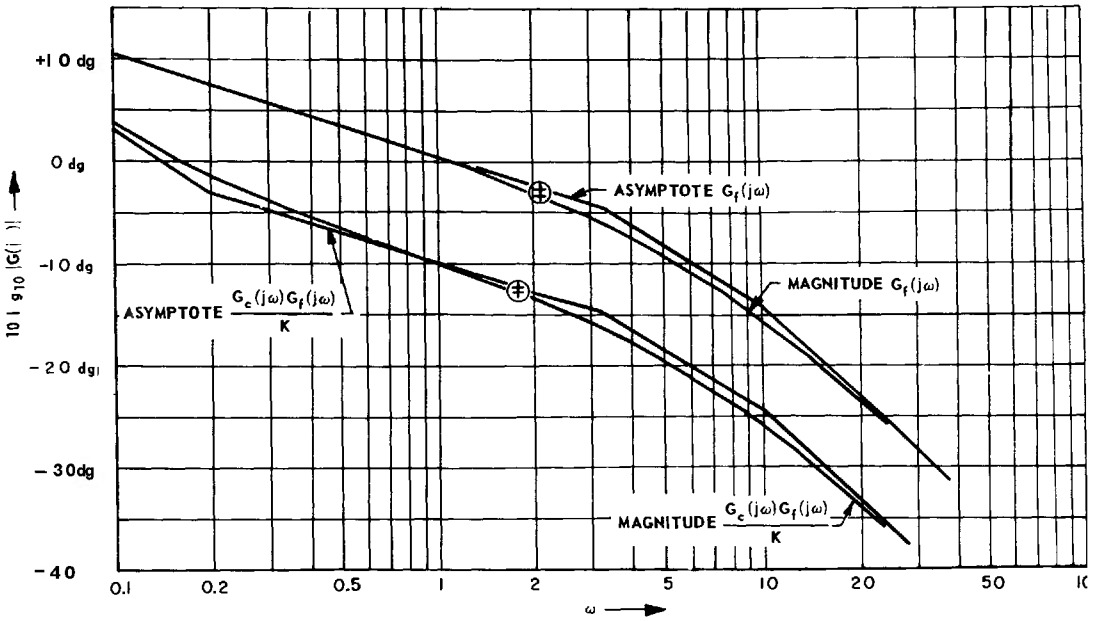


Fig. 6-7 Magnitude curves for lag-compensation procedure employing phase margin.

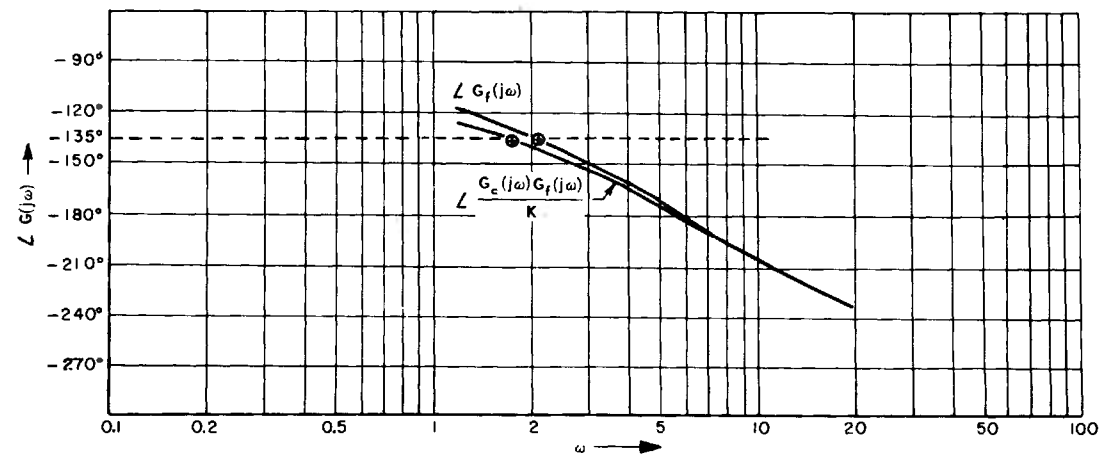


Fig. 6-8 Phase curves for lag-compensation procedure employing phase margin.

(b) The magnitude and asymptote curves of $G_f(j\omega)$ are drawn in Fig. 6-9 and the phase angle curve is drawn in Fig. 6-10. Using the problem specifications and Eqs. (6-11) and (6-14), $\psi = 45°$, $\phi_m = 54.9°$ and $\beta_m = 0.316$. Eq. (6-12) yields

$$\angle\, G_f(j\omega_1) = -189.9° \qquad (6\text{-}21)$$

Using this result and the $G_f(j\omega)$ phase-angle curve (Fig. 6-10), $\omega_1 = 7.2$ rad/sec. From Eq. (6-19), the initial-trial time constant $T_{c_1} = 0.044$ sec and the scale change from β to ω is given by

$$\beta = 0.044\omega \qquad (6\text{-}22)$$

The composite $G_c(j\omega)G_f(j\omega)/K$ magnitude and asymptote curves for the first-trial lead-compensated system are drawn in Fig. 6-9

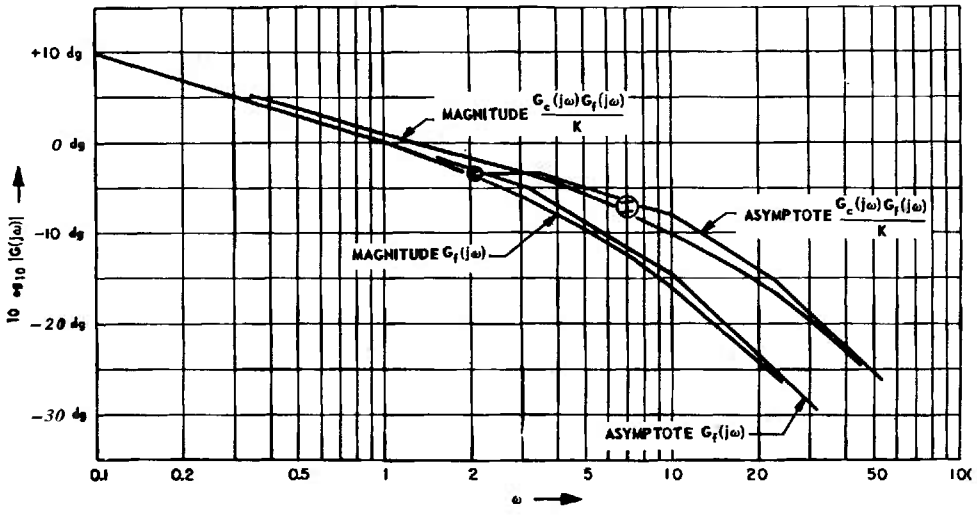


Fig. 6-9 Magnitude curves for lead-compensation procedure employing phase margin.

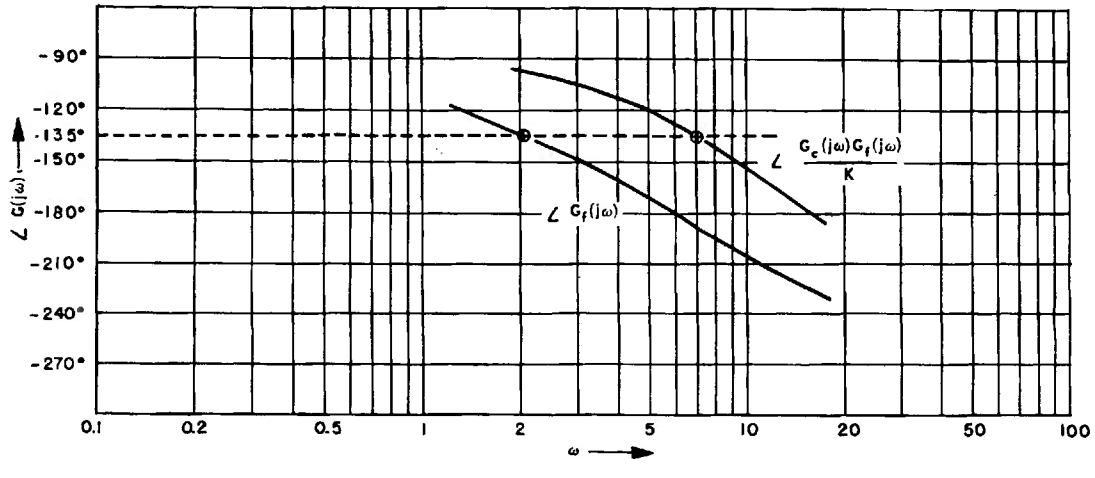


Fig. 6-10 Phase curves for lead-compensation procedure employing phase margin.

and the corresponding phase-angle curve is drawn in Fig. 6-10. Using the 45° phase-margin criterion for the lead compensated system, the magnitude-crossover frequency $\omega_{cm} = 7.0$ rad/sec, and the corresponding gain is 6.02. The asymptote-crossover frequency $\omega_c = 7.0$ rad/sec and the corresponding gain is 4.46. The results of this example are summarized in Table 6-2.

TABLE 6-2 RESULTS OF COMPENSATION USING 45° PHASE MARGIN

Compensation Adjustment	ω_{cm}	K for ω_{cm}	ω_c	K for ω_c	Factor of Gain Increase	Factor of ω_c or ω_{cm} Change
No compensation; true magnitude used	2.08	2.34				
No compensation; asymptote used			2.08	2.04		
Lag compensation; true magnitude used	1.74	20			8.6	0.84
Lag compensation; asymptote used			1.74	17.8	8.7	0.84
Lead compensation; true magnitude used	7.0	6.02			2.6	3.4
Lead compensation; asymptote used			7.0	4.46	2.2	3.4

6-4 FEEDBACK OR PARALLEL COMPENSATION [2,3,4]

The cascade type of compensation discussed in Pars. 6-2 and 6-3 has the disadvantage that the compensation adjustment is sensitive to changes in the parameters of the fixed elements due to non-linear behavior of the system. When feedback compensation is used, on the other hand, the compensation adjustment is much less sensitive to fixed-element parameter variation provided the loop gain is high. In addition, the networks used in feedback compensation are usually simpler in form than the corresponding cascade networks. However, the necessity for high loop gain (at least a gain of 10 at the break frequencies of the feedback networks) generally requires a more complicated and expensive system.

The procedure used in designing feedback compensation networks employs a combination of the gain-phase plane and the asymptotic-magnitude presentations. The basic principles involved in the design of feedback compensation networks can be clarified by a study of Fig. 6-11. Here a feedback function $H_c(s)$ is used to modify the characteristics of the fixed elements $G_f(s)$ and a cascade function $G_c(s)$ is provided to aid in adjusting the

performance of the major loop. In most cases, the cascade function $G_c(s)$ is a pure gain K which serves to adjust the degree of stability of the system. The burden of reshaping the $G_f(s)$ function is placed on the feedback compensation function $H_c(s)$. In addition, the feedback function is usually provided with a gain adjustment to permit the setting of the degree of stability of the minor loop.

The general configuration of Fig. **6-11** can be redrawn and placed in the cascade form of Fig. **6-12.** Here,

$$G'_c(s) = \frac{G_c(s)}{1 + G_f(s)H_c(s)} \qquad \textbf{(6-23)}$$

Thus, theoretically, cascade compensation and feedback compensation are equivalent. In practice, feedback compensation is more flexible, and the resulting system is less sensitive to component parameter variations.

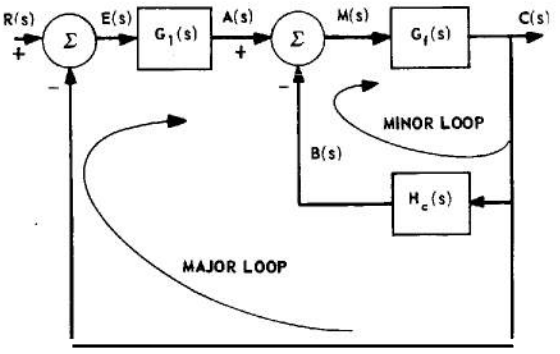


Fig. 6-11 General feedback-compensation configuration.

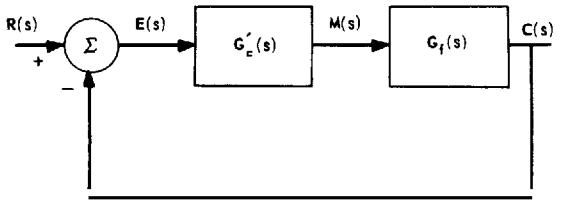


Fig. 6-12 Cascade equivalent of feedback-compensation configuration.

The procedure for adjusting the feedback compensation is best understood by examining the asymptotic behavior of the minor loop. If

$$|G_f(j\omega)H_c(j\omega)| >> 1,$$

then

$$\frac{C(j\omega)}{A(j\omega)} = \frac{G_f(j\omega)}{1 + G_f(j\omega)H_c(j\omega)} \cong \frac{1}{H_c(j\omega)} \qquad \textbf{(6-24)}$$

If

$$|G_f(j\omega)H_c(j\omega)| << 1,$$

then

$$\frac{C(j\omega)}{A(j\omega)} \cong G_f(j\omega) \qquad \textbf{(6-25)}$$

Thus, in the frequency ranges where the open-minor-loop frequency-response magnitude $|G_f(j\omega)H_c(j\omega)|$ is very large, the closed-minor-loop response $C(j\omega)/A(j\omega)$ behaves like the reciprocal of the feedback function $H_c(j\omega)$. When the open-minor-loop-response magnitude is very small, the closed-minor-loop response behaves like the fixed elements response function $G_f(s)$. Thus, the frequency scale can be divided into several regions based on the magnitude of the open-minor-loop response. Whenever the magnitude of this response $|G_f(j\omega)H_c(j\omega)|$ or $|B(j\omega)/M(j\omega)|$ is greater than unity, the asymptotes of the closed-minor-loop response coincide with the asymptotes of the reciprocal of the feedback function $H_c(j\omega)$. Whenever the magnitude of the $B(j\omega)/M(j\omega)$ function is less than unity, the closed-minor-loop response asymptotes coincide with the asymptotes of the fixed-element response $G_f(j\omega)$.

Usually, the feedback function $H_c(j\omega)$ is so chosen that the frequency scale is divided into three ranges. These ranges are:

$$0 < \omega < \omega_l \quad \text{when } |G_f(j\omega)H_c(j\omega)| < 1 \qquad \textbf{(6-26)}$$

$$\omega_l < \omega < \omega_u \quad \text{when } |G_f(j\omega)H_c(j\omega)| > 1 \qquad \textbf{(6-27)}$$

$$\omega_u < \omega < \infty \quad \text{when } |G_f(j\omega)H_c(j\omega)| < 1 \qquad \textbf{(6-28)}$$

and ω_l is the lower-frequency boundary and ω_u is the upper-frequency boundary. Corresponding to the three frequency ranges fixed by the magnitude of the open-minor-loop response, the closed-minor-loop response asymptotes are defined by

$$\left|\frac{C(j\omega)}{A(j\omega)}\right| \cong \left|G_f(j\omega)\right| \quad \text{for } 0 < \omega < \omega_l \tag{6-29}$$

$$\left|\frac{C(j\omega)}{A(j\omega)}\right| \cong \left|\frac{1}{H_c(j\omega)}\right| \quad \text{for } \omega_l < \omega < \omega_u \tag{6-30}$$

$$\left|\frac{C(j\omega)}{A(j\omega)}\right| \cong \left|G_f(j\omega)\right| \quad \text{for } \omega_u < \omega < \infty \tag{6-31}$$

The procedure for adjusting the feedback function $H_c(s)$ and the cascade gain $G_c(s) = K$ can thus be roughed out using asymptotic pictures of the various responses and then carried out in detail using the gain-phase plane. The aim of the asymptotic sketches is to examine the form of the closed-minor-loop response as the feedback compensation is adjusted. Examination of the asymptotes of typical cascade compensation arrangements can also serve as a guide to the shaping of the closed-minor-loop response. The desirable properties of the closed-minor-loop response can be expressed in terms of the properties that are desirable for any open-major-loop function; namely, high gain at low frequencies ($\omega < \omega_l$), a stable shape relative to the $-1 + j0$ point, and low gain at high frequencies ($\omega > \omega_u$).

Since there are usually several parameters to adjust in the feedback compensation procedure, the process of adjustment is one of trial and error guided by the asymptotic sketches. The details of the procedure are best demonstrated by an example.

Example. Th ransfer function of the fixed elements of a system is given by

$$G_f(s) = \frac{1}{s(0.3s + 1)(0.1s + 1)} \tag{6-32}$$

Feedback compensation is to be used to improve the performance of the system in conjunction with a pure gain cascaded with the minor loop. The transfer function of the feedback elements is given by

$$H_c(s) = \frac{K_c s^2}{T_c s + 1} \tag{6-33}$$

Note that this transfer function can be realized in a position-control system by a tachometer cascaded with a single-stage high-pass RC filter. The cascade compensation being a pure gain, its transfer function is given by

$$C_c(s) = K \tag{6-34}$$

Design the compensation for a 45° major-loop phase margin.

Solution. The open-minor-loop transfer function is

$$\frac{B(s)}{M(s)} = G_f(s)H_c(s) = \frac{K_c s}{(T_c s + 1)(0.3s + 1)(0.1s + 1)} \tag{6-35}$$

A plot of the asymptotes of $G_f(j\omega)H_c(j\omega)/K_c$, using the techniques for plotting asymptotic magnitude curves of Par. 5-3.3, is shown in Fig. 6-13 for $T_c = 2$ sec as a trial guess. The adjustment of K_c controls the degree of stability of the closed-minor-loop response. If K_c is made too large (e.g., greater than 50), the closed-minor-loop response will have a quadratic factor with a very low damping ratio, making it difficult to obtain a high-gain open-major-loop response. Anticipating this behavior, a value of $K_c = 10$ is not unreasonable. With $K_c = 10$, the inequality of Eq. (6-27) is satisfied for the portions of the open-minor-loop asymptotes above the −10 dg line. When K_c is set equal to 10, the 0-dg line for the open-minor-loop response $G_f(j\omega)H_c(j\omega)$ is that shown dashed in Fig. 6-13. This line defines the frequency boundaries $\omega_l = 0.1$ rad/sec and $\omega_u = 13$ rad/sec. The closed-minor-loop response asymptotes can then be drawn with the aid of the approximations to $|C(j\omega)/A(j\omega)|$ given in Eqs. (6-29) to (6-31). In the frequency range from ω_l to $\omega = 1/T_c$ there is a −20 dg/dec contribution to $|1/H_c(j\omega)|$ from

the factor s^2 whereas the $(T_c s + 1)$ factor makes no contribution (i.e., it contributes 0 dg/dec). In the range from $\omega = 1/T_c$ to ω_u, both factors contribute, the s^2 factor contributing −20 dg/dec and the $(T_c s + 1)$ factor + 10 dg/dec. The closed-minor-loop asymptotes are shown in Fig. **6-14.** From an examination of the resultant asymptotic curve, several points may be noted. The combination of the breaks at w_l and $1/T_c$ appears as a cascade lag-compensation effect which is a desirable open-major-loop-response property. In further trials, attempts may be made to broaden the −20 dg/dec slope region bounded by these breaks and to move the region to a higher frequency range. The break at ω_u is from a slope of −10 dg/dec to a slope of −30 dg/dec which is characteristic of a quadratic factor in the open-major-loop response. If this factor has a low damping ratio, high-gain stabilization of the major loop may be difficult. Thus, the first trial choices of T_c and K_c produce a set of open-major-loop asymptotes which appear reasonable; however the adequacy of the choices must be verified.

At this point, the Nichols chart (Fig. **5-20)** is used with a gain-phase plot of $G_f(j\omega)H_c(j\omega)$ to determine the closed-minor-loop response $C(j\omega)/A(j\omega)$ (see Par. **5-4** for the use of the Nichols chart with non-unity-feedback loops).

As a result of the application of the gain-phase plane construction, the true magnitude curve of $C(j\omega)/A(j\omega)$ is shown in Fig. **6-14** can be obtained. The corresponding phase angle curve appears in Fig. **6-15.** The shape of the true magnitude curve shows no severe resonance effects so that acceptable closed-loop performance may be expected. Using the 45° phase-margin criterion to adjust the cascade compensation $G_c(s) = K$, the magnitude-crossover frequency $\omega_{cm} = 8.6$ rad/sec and the corresponding gain $K = 40$. Thus, the performance is quite good. (Compare the results of cascade compensation for this same system in Pars. **6-2** and **6-3.)** The only drawback to the design is that the equivalent cascade lag effect is at a fairly low frequency. This would produce somewhat excessive peaking in the transient response of the system which would be followed by a long transient tail. Improvement in performance could be achieved by further trial, e.g., by decreasing the feedback compensation time constant T_f and attempting to increase the minor-loop gain K_c. The resultant system would then have a more acceptable transient behavior, but the magnitude-crossover frequency ω_{cm} and the major-loop gain K might be reduced. However, only further trial-and-error analysis would show what actually occurs.

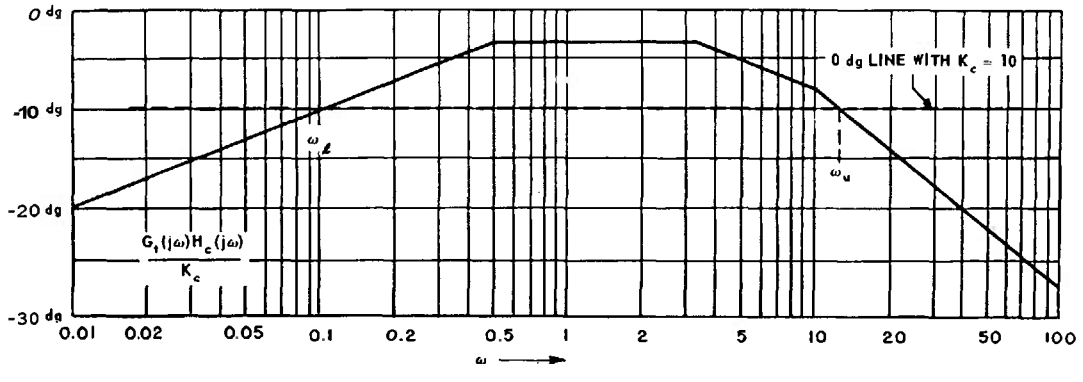


Fig. 6-13 Open-minor-loop asymptote for feedback compensation procedure.

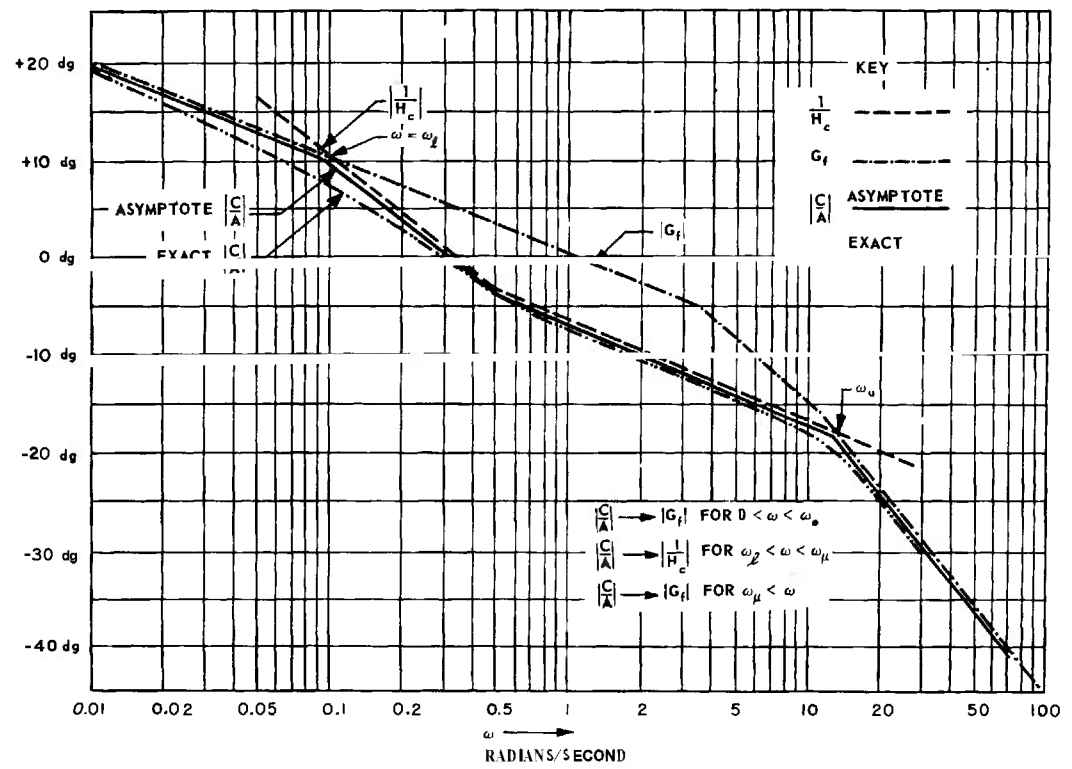


fig. 6-14 Closed-minor-loop magnitude for feedback compensation procedure.

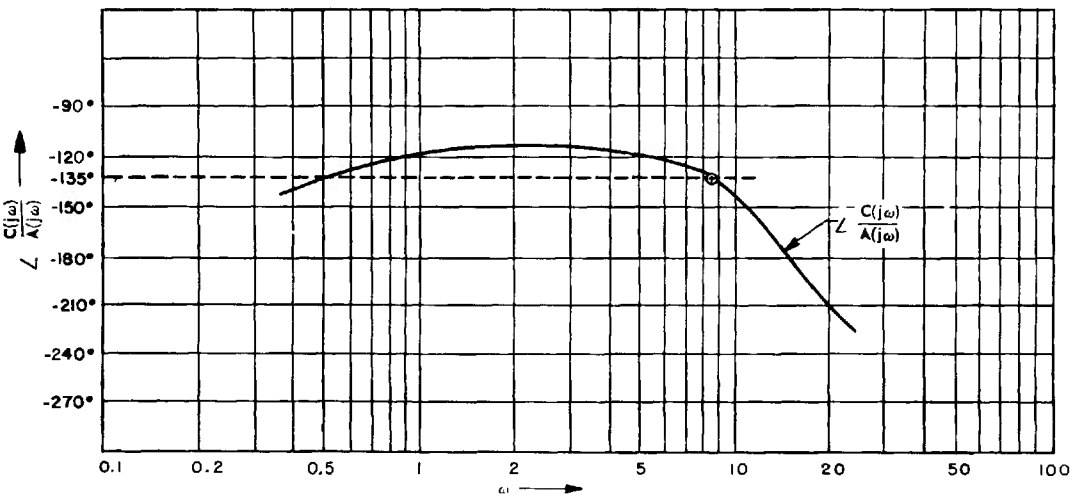


Fig. 6-15 Closed-minor-loop phase angle for feedback compensation procedure.

6-5 ALTERNATIVE DESIGN METHODS[6,7,9,10,12,13]

The primary advantage of design in the frequency domain is the rapidity with which the procedure can be carried out. The disadvantage is the difficulty involved in visualizing the time-domain behavior corresponding to a given frequency-domain design. In practice, the relationship between frequency response and transient response is considered by many workers to be rather nebulous. Theoretically, however, frequency-domain and time-domain descriptions are entirely equivalent, although the actual process of translating from one description to the other may be quite laborious in spite of the fact that many approximations have been established to guide the designer in relating frequency response with transient response (see Par. 7-1). One very important stumbling block arises because most direct specifications of system performance are given in terms of the transient response of the system to a step or ramp input. This type of specification is just as artificial as that given in terms of the response of a system to a sinusoidal input since the true inputs of most systems are neither steps, nor ramps, nor sinusoids. Nevertheless, it is the transient response of a system that is most frequently specified because this type of response is the easiest to visualize and the quickest to verify experimentally.

In order to circumvent some of the conceptual difficulties involved in frequency-domain design, methods of time-domain design have been advanced. Most of these methods utilize the open-loop and closed-loop pole-zero configurations of the system and thus involve features of both the time and frequency domains. The facility with which these methods can be used depends almost exclusively on the feature of having an analytical description of the open-loop pole-zero configuration as a starting point. Thus, the methods require that any experimental test data be approximated by analytical functions. This requirement does not apply to the frequency-domain methods that receive major emphasis in Ch. 5 and Pars. 6-2 through 6-4. In addition, the graphical procedures discussed in Pars. 3-6 and 7-1 enable the designer to work entirely with experimental data, going back and forth between time and frequency domains without ever having to deal with analytical descriptions. Since the time-domain synthesis methods usually end up with a closed-loop pole-zero configuration, additional labor is necessary to extract the actual plots of transient response and frequency response in order to verify whether performance specifications have been met. On the other hand, in a frequency-domain design, the only additional labor involved is that of determining the transient response (usually by the methods of Par. 3-6), the frequency response being directly available at the end of the design procedure. Thus far, the time-domain procedures that have been developed are most successful for analysis but are quite time-consuming and laborious for synthesis. Actually, most of the current time-domain "synthesis" procedures merely involve ordered trial-and-error analysis. A few of the time-domain methods are described here.

Evan's root-locus method[6,7] can be used for the design of compensation functions by postulating a series of trial forms of the proposed compensation functions and plotting the root locus for each form (see Pars. 4-4 and 5-7 for the technique of root-locus construction and the nature of the degree-of-stability criterion). On adjusting the gain to satisfy the degree-of-stability criterion with a specified damping ratio ζ for the dominant pole pair, each trial root locus will produce a specific closed-loop pole-zero configuration. Then by direct inspection of these configurations or by plots of the actual transient responses (through partial-fraction expansion and inverse Laplace transformation), the best compensation form may be selected.

Yeh[13] has proposed an extension of Evan's method which involves plotting contours of closed-loop pole location for a series of *fixed-gain* values as some parameter of the compensation function is varied. These plots are called

gain contours. In addition, for *fixed* values of the compensation-function parameter, contours of closed-loop pole location are plotted as the gain is varied. These plots are called *root contours*. By examining the gain and root contours, the best combination of the compensation-function parameter and the gain can be selected.

Truxal[6] has developed a pure synthesis procedure based on the desired closed-loop pole-zero configuration (see Par. 7-1). It is assumed that this configuration is characterized by: (1) one pair of conjugate-complex dominant poles, (2) one or more dipoles (a pole and zero very close together on the negative real axis), (3) poles on the negative real axis that are far removed from the dominant pole pairs, and (4) one or more finite zeros[9]. This closed-loop pole-zero configuration will be produced by an open-loop function [$C(s)/E(s)$ for a unity feedback system] which has all its poles on the negative real axis. If the closed-loop function

$$\frac{C(s)}{R(s)} = \frac{P(s)}{N(s)} \qquad (6\text{-}36)$$

and, for a unity-feedback system,

$$\frac{C(s)}{E(s)} = \frac{P(s)}{Q(s)} \qquad (6\text{-}37)$$

then

$$Q(s) = N(s) - P(s) \qquad (6\text{-}38)$$

where $P(s)$, $N(s)$, and $Q(s)$ are polynomials in s.

The synthesis procedure is then a method for determining the zeros of Eq. (6-38) since these are the poles of $C(s)/E(s)$. Since all the poles of $C(s)/E(s)$ lie on the negative real axis, if the polynomials $N(s)$ and $P(s)$ are plotted on the same coordinate system for $s = -\sigma$ where σ is a real variable, then the intersections of the two curves give the poles of $C(s)/E(s)$. The zeros of $C(s)/E(s)$ are the same as the zeros of $C(s)/R(s)$. Knowing the transfer function of the fixed elements $G_f(s)$, the compensation can be determined from the following equation:

$$G_c(s) = \frac{1}{G_f(s)}\left[\frac{C(s)}{E(s)}\right] \qquad (6\text{-}39)$$

The cancellation of the function $G_f(s)$ by the compensation $G_c(s)$ should be avoided as much as possible by having some of the poles of $G_f(s)$ occur in the open-loop function $C(s)/E(s)$. This can be accomplished by altering slightly the specified form of the closed-loop response $C(s)/R(s)$ since the performance specifications are rarely rigid. Changes in the parameters of the fixed elements $G_f(s)$ will negate the cancellation called for by Eq. (6-39). Actually, exact cancellation is not necessary since small parameter variations will not alter the closed-loop response appreciably.

6-6 TYPICAL COMPENSATION NETWORKS[2,5,6,14,15,16,17,18,19,20,21]

6-6.1 D-C ELECTRIC

The most common d-c networks are the lag network and the lead network shown in Fig. 6-16.

The ***lag-network*** transfer function is

$$\frac{E_o(s)}{E_i(s)} = \frac{Ts + 1}{\alpha Ts + 1} \qquad (6\text{-}40)$$

where

$$T = R_2C, \text{ and} \qquad (6\text{-}41)$$

$$\alpha = 1 + \frac{R_1}{R_2} \qquad (6\text{-}42)$$

The *lead network* transfer function is

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{a}\,\frac{\alpha Ts + 1}{Ts + 1} \qquad (6\text{-}43)$$

where

$$T = \left(\frac{R_1R_2}{R_1 + R_2}\right)C, \text{ and} \qquad (6\text{-}44)$$

$$\alpha = 1 + \frac{R_1}{R_2} \qquad (6\text{-}45)$$

Chestnut and Mayer[5] present a series of charts of d-c networks containing only resistances and capacitances. These charts cover most of the desirable frequency-response characteristics that are called for in compensation of feedback control systems.

6-6.2 A-C ELECTRIC[2, 5, 6, 14, 15, 16, 17, 18, 19, 20]

In many control system applications, the signals are suppressed-carrier modulated, and the control information modulates a constant-amplitude carrier signal (in practice, 60 or 400 cps). For example, the form of voltage corresponding to the actuating signal may be as follows:

$$V(t) = e(t) \cos \omega_o t$$

where $e(t)$ is the true data signal, and ω_o is the carrier frequency. Networks which are designed to operate on the data of carrier-modulated signals are called a-c or carrier-frequency networks. If it is necessary to compensate a system employing carrier-modulated signals a-c networks are required since d-c networks will not work because they effectively operate on zero-frequency-carrier signals.

There are two questions involved in treating the compensation of carrier-modulated signals:

(1) Analysis: Given a network which operates on a carrier-modulated signal, what is the data-frequency (d-c) equivalent network?

(2) Synthesis: Given a desired data-frequency (d-c) network, what is the equivalent carrier-frequency network?

If $H(j\omega)$ is the frequency response of a carrier network, the frequency response of the data-frequency equivalent is given by

$$H(j\omega_d) = \frac{1}{2}\sqrt{A^2 + B^2}\, e^{j\psi_d} \qquad (6\text{-}46)$$

where

$$A = |H_+| \cos(\angle H_+) + |H_-| \cos(\angle H_-) \qquad (6\text{-}47)$$

$$B = -|H_+| \sin(\angle H_+) + |H_-| \sin(\angle H_-) \qquad (6\text{-}48)$$

$$\psi_d = \tan^{-1}\frac{B}{A} \qquad (6\text{-}49)$$

$$H_+ \triangleq H[j(\omega_o + \omega_d)] \qquad (6\text{-}50)$$

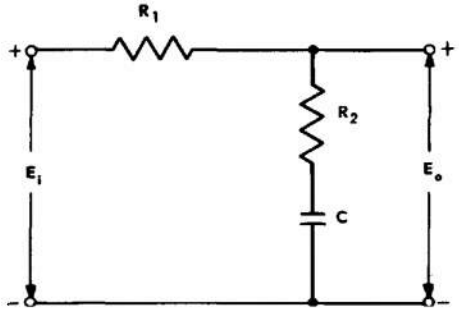
$$H_- \triangleq H[j(\omega_o - \omega_d)] \qquad (6\text{-}51)$$

ω_o = carrier freyuency

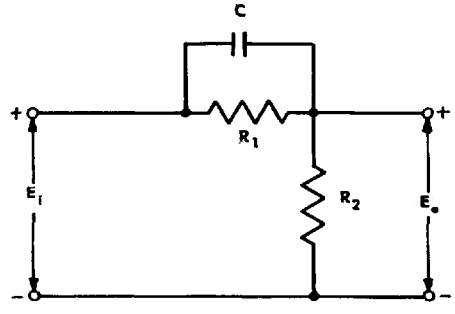
ω_d = data frequency

$H(j\omega_d)$ = frequency response of equivalent data-frequency network.

There is no unique answer to the synthesis problem, but a convenient answer is given by the "low-pass to band-pass" transformation. If it is assumed that the magnitude of the carrier-frequency equivalent of a data-frequency network has even symmetry about the



A. LAG NETWORK



B. LEAD NETWORK

Fig. 6-76 D-C compensation networks.

carrier frequency ω_o and that the phase angle has odd symmetry about ω_o, i.e.,

$$|H_+| = |H_-| \tag{6-52}$$

$$\angle H_+ = -\angle H_-, \tag{6-53}$$

then the carrier-frequency equivalent to a given data-frequency network is

$$H(j\omega) = H[j(\omega - \omega_o)] \tag{6-54}$$

Unfortunately, an exact solution of Eq. (6-54) leads to the conclusion that if the data-frequency network H is physically realizable, the carrier-frequency network H is not. However, an approximation to the low-pass to band-pass transformation is possible which does lead to physically realizable networks. If the data frequencies ω_d are small compared to the carrier frequency ω_o, then

$$j(\omega - \omega_o) \cong j\frac{\omega}{2} + \frac{\omega_o^2}{2j\omega} \tag{6-55}$$

As an example, suppose that the frequency response of a data-frequency network is given by

$$H(j\omega_d) = \frac{T_n j\omega_d + 1}{T_d j\omega_d + 1} \tag{6-56}$$

The carrier-frequency equivalent can be found by using Eqs. (6-54) and (6-55). Thus

$$H(j\omega) \cong \frac{T_n\left[j\frac{\omega}{2} + \frac{\omega_o^2}{2j\omega}\right] + 1}{T_d\left[j\frac{\omega}{2} + \frac{\omega_o^2}{2j\omega}\right] + 1} \tag{6-57}$$

There are several ways to realize a carrier-frequency network which is approximately equivalent to a given data-frequency network. The resistance-inductance-capacitance realization starts with the actual data-frequency circuit and replaces the data-frequency circuit elements by their approximate carrier-frequency equivalents as shown in Fig. 6-17. Because of the practical difficulty of realizing parallel inductance-capacitance combinations in the carrier-frequency network, the usual procedure is to realize the data-frequency transfer function by means of a resistance-inductance circuit. Then the carrier equivalent will contain only series inductance-capacitance combinations.

Because lag networks are usually inserted at very low data frequencies, their carrier equivalents are required to be very sharply tuned to the carrier frequency. That is, the carrier equivalent network must be a high-Q circuit. Unfortunately, such high-Q circuits are impractical for servo carrier frequencies (60 and 400 cps) and are very sensitive to

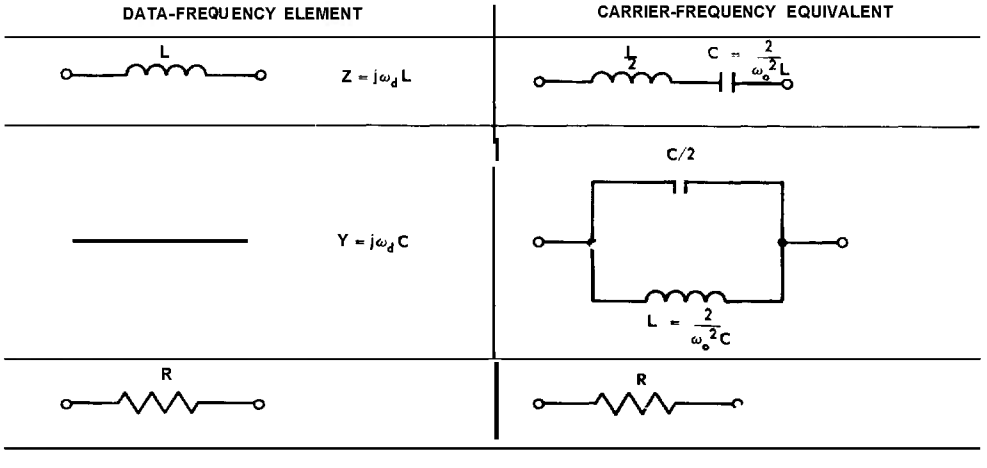


Fig. 6-17 Equivalent circuit elements for carrier-frequency networks.

carrier frequency drift. To get around these limitations when lag compensation is desired for a carrier-modulated system, the usual procedure is to demodulate the signal, compensate it with a d-c lag network, and then remodulate.

The realization of carrier-frequency lead networks is not as difficult as the realization of lag networks since they operate at relatively high data frequencies and therefore do not require excessively high-Q circuits. However, carrier lead networks are also sensitive to carrier drift, although some attempts have been made to counteract this effect.[15,18] If the carrier drift is large (more than 5%), then the scheme of demodulation, compensation, and remodulation should be considered. An effective alternative to this scheme utilizes feedback compensation with a tachometer.

An examination of the magnitude responses of carrier networks shows that they fall into the class of filters called "notch" filters. Methods for realizing notch filters with resistance-capacitance rather than resistance-inductance-capacitance networks are discussed in Refs. (2,5,14). A typical resistance-capacitance notch filter is shown in Fig. 6-18. The frequency response of this carrier network is given by

$$\frac{E_o(j\omega)}{E_i(j\omega)} = H(j\omega) = \frac{\left(j\frac{\omega}{\omega_o}\right)^2 + 1}{\left(j\frac{\omega}{\omega_o}\right)^2 + 4\left(j\frac{\omega}{\omega_o}\right) + 1} \qquad (6\text{-}58)$$

where

$$\omega_n = \frac{1}{RC} \qquad (6\text{-}59)$$

The approximate data-frequency equivalent is

$$H(j\omega_d) \cong \frac{j\omega_d}{j\omega_d + 2\omega_o} \qquad (6\text{-}60)$$

Thus, as far as data frequencies are concerned, the symmetrical parallel $-T$ notch filter behaves as a differentiator for data frequencies up to approximately $0.2\omega_o$.

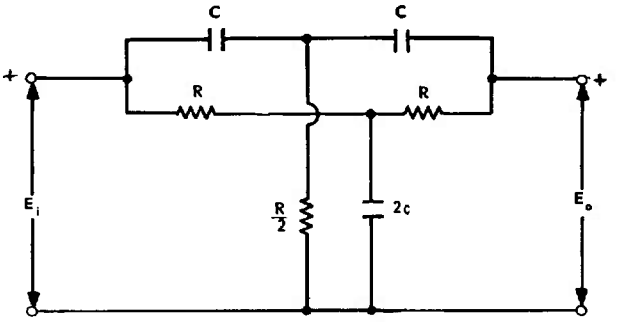


Fig. **6-78** *Parallel-T* notch filter.

The major difficulty in using resistance-capacitance notch filters is that they must be tuned by successive adjustments of several circuit elements; otherwise, high-precision elements must be used.

6-6.3 MECHANICAL DAMPER

A widely used mechanism having the action of a compensation network is the ***inertia damper*** shown in Fig. 6-19. The damper, which is connected directly to the servomotor shaft, consists of a thin cylindrical metal shell, a heavy cylindrical metal slug, and a damping fluid. If one neglects motor damping, the block diagram of the inertia damper and servomotor is that shown in Fig. 6-20. In this figure,

$$T = \frac{J_d(J_s + J_m)}{B(J_s + J_m + J_d)} \qquad (6\text{-}61)$$

$$\alpha = 1 + \frac{J_d}{J_s + J_m} \qquad (6\text{-}62)$$

T_m = motor torque

T_L = load torque

θ_m = motor shaft position

$J_{,,,}$ = motor moment of inertia and reflected load inertia

J_s = shell moment of inertia

J_d = slug moment of inertia

B = fluid damping

The advantages of the inertia damper are:

(1) simplicity

(2) no steady power loss

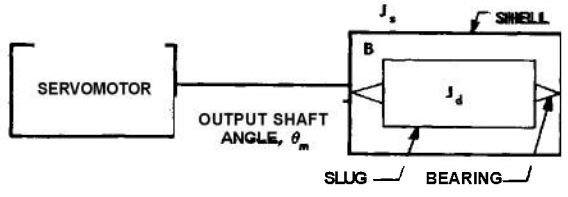


Fig. 6-79 Inertia damper.

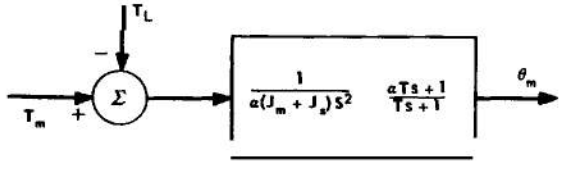


Fig. 6-20 Block diagram of inertia damper (motor damping negligible).

The disadvantages are :

(1) damper must be designed and built for each specific application.

(2) peak acceleration **of** the damper-motor combination is reduced relative to that **of** the motor alone because **of** the added apparent inertia produced at the motor shaft by the damper mechanism.

6-6.4 HYDRAULIC AMPLIFIER (See Par.13-6)

A fairly common means **for** obtaining *lag-network* action in a hydraulic amplifier is shown in Fig. **6-21.** In this figure,

x_i = input displacement of pilot valve
x_o = output motion **of** power piston
x_s = feedback motion **of** follow-up sleeve
B = damping **of** fluid dashpot
K_1, K_2 = spring constants
a, b = lever arms

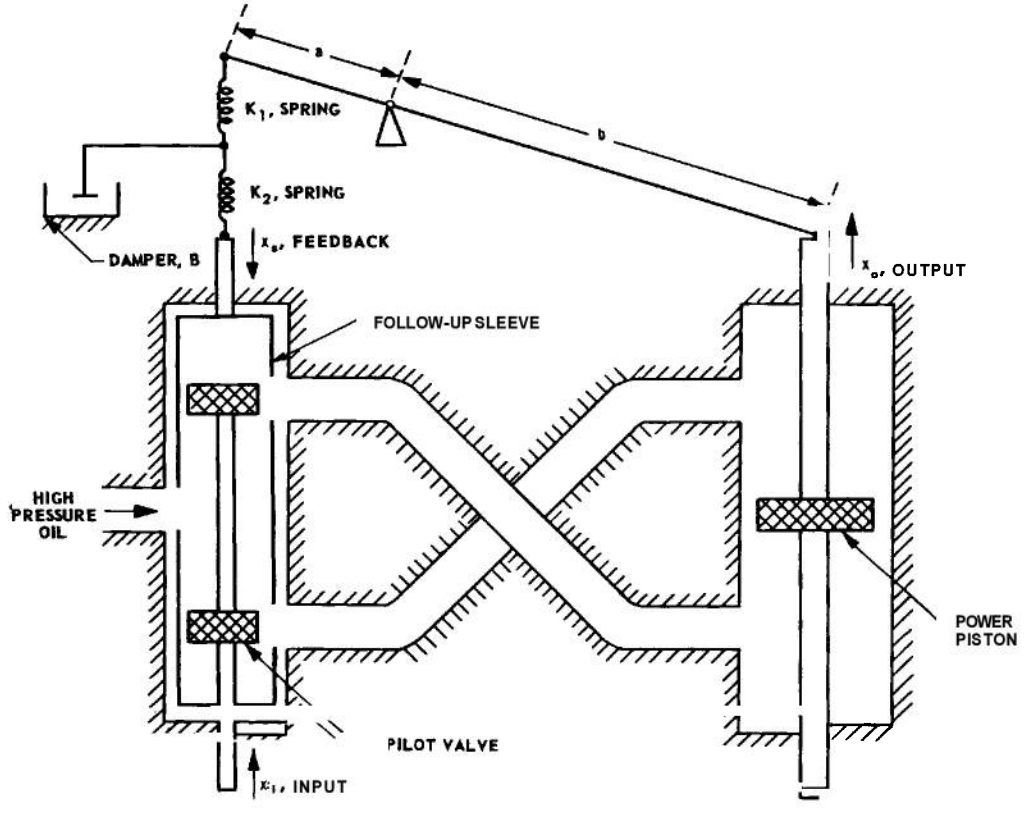


Fig. 6-21 Hydro-mechanical compensation network.

If the gain of the hydraulic amplifier is very high (greater than 10), the transfer function of the system is

$$\frac{x_o(s)}{x_i(s)} \cong \frac{b}{a}\alpha(\alpha - 1)\left(\frac{1 + Ts}{1 + \alpha Ts}\right) \qquad (6\text{-}63)$$

where

$$\alpha = 1 + \frac{K_2}{K_1} \qquad (6\text{-}64)$$

and

$$T = B/(K_2 + K_1) \qquad (6\text{-}65)$$

6-6.5 PNEUMATIC CONTROLLER[21] (See Par. 13-7.)

The general schematic of a typical pneumatic controller is shown in Fig. 6-22. In this figure,

r = motion of set point (reference input)

c = motion of pen (controlled variable or output)

e = actuating signal

x_f = flapper motion

P_r = nozzle back pressure

P_o = pilot relay output pressure (to diaphragm valve or similar load)

x_{fb} = feedback motion.

If the nozzle-flapper amplifier and the pilot relay are assumed ideal, the block diagram of the controller is as shown in Fig. 6-23, where

K_1 = ratio of proportioning linkage $(0 < K_1 < 1)$

k_{nf} = nozzle-flapper gain

k_{pr} = pilot relay gain

Simple *proportional action* (pure gain) is possible if the feedback function is achieved by means of a spring-loaded bellows as shown in Fig. 6-24. If the ratio of feedback motion x_{fb} to pilot-relay output pressure P_o is denoted

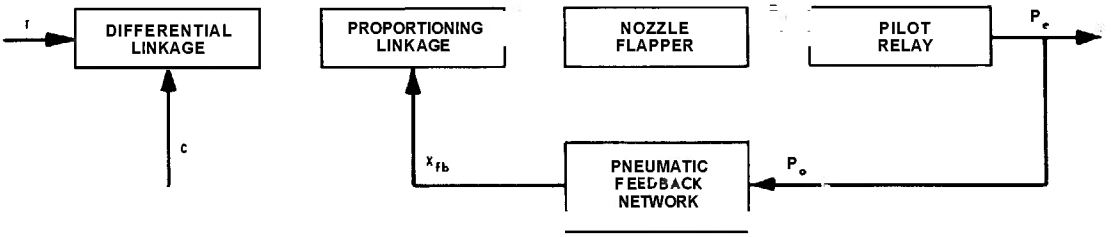


Fig. 6-22 Schematic diagram of a pneumatic controller.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

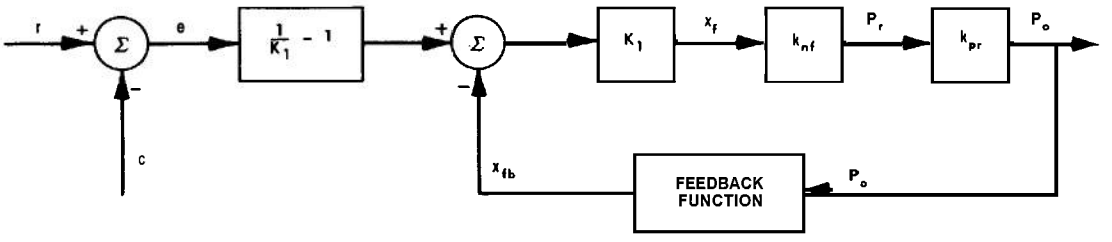


Fig. 6-23 Controller block diagram.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1963, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

by K_{fb}, the transfer function of the proportional controller is given by

$$\frac{P_o(s)}{e(s)} = \frac{(1 - K_1)k_{nf}k_{pr}}{1 + K_1K_{fb}k_{nf}k_{pr}} \qquad (6\text{-}66)$$

If the product $K_1K_{fb}k_{nf}k_{pr}$ is very high (greater than 10), the proportional controller has the approximate response

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1K_{fb}} \qquad (6\text{-}67)$$

A *lag-compensation effect* (proportional-plus-integral) can be achieved if the feedback function of Fig. 6-22 is obtained by means of the arrangement of Fig. 6-25. The feedback function in this case is given by

$$\frac{x_{fb}(s)}{P_o(s)} = K_{fb}\left(\frac{T_Rs}{T_Rs + 1}\right) \qquad (6\text{-}68)$$

where

$T_R \triangleq R_RC_R$

R_R = integral resistance

C_R = capacitance of tank

K_{fb} = sensitivity of proportional bellows

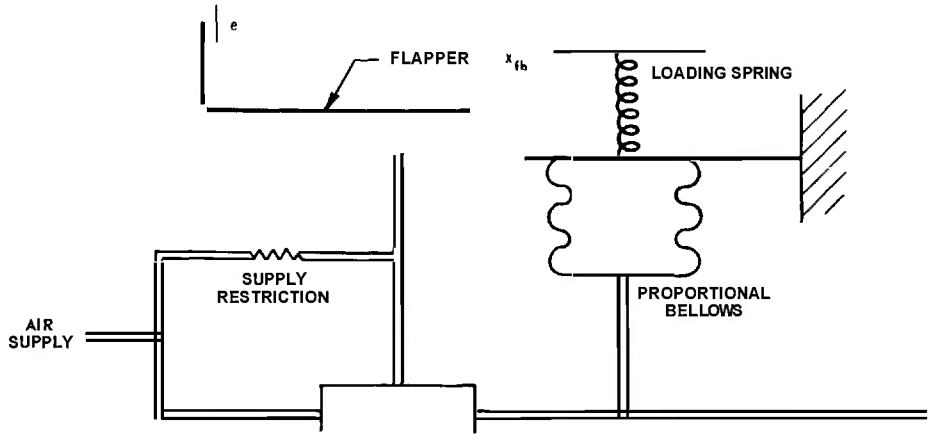


Fig. 6-24 Schematic diagram of a proportional controller.

Adapted by permission from *Instruments*. Volume 26, No. 6. June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Could and P. E. Smith, Jr.

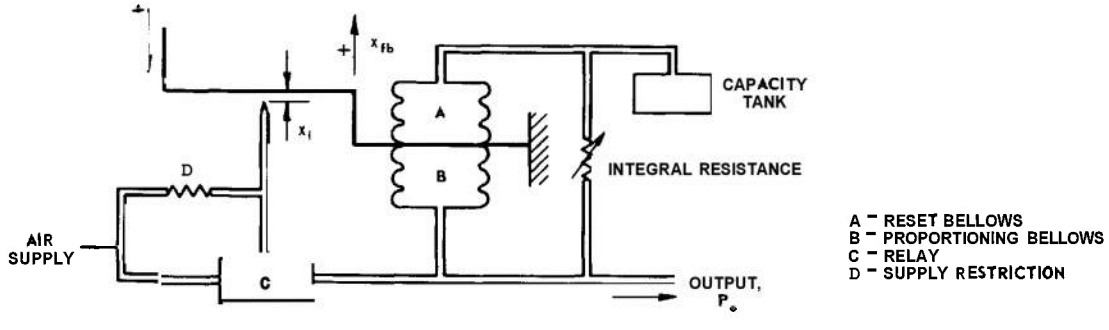


Fig. 6-25 Schematic diagram of a proportional plus integral controller.

Adapted by permission from *Instruments*. Volume 26, No. 6, June, 1953. from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

The transfer function of the controller then becomes

$$\frac{P_o(s)}{e(s)} = (1 - K_1)\, k_{nf} k_{pr} \left(\frac{T_R s + 1}{\alpha_R T_R s + 1} \right) \tag{6-69}$$

where

$$\alpha_R = 1 + K_1 k_{nf} k_{pr} K_{fb} \tag{6-70}$$

If the product $K_1 k_{nf} k_{pr} K_{fb}$ is very high (greater than 50), the response of the proportional-plus-integral controller is approximately

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} \left(1 + \frac{1}{T_R s} \right) \tag{6-71}$$

The form of the right side of this equation explains the name — "proportional-plus-integral" controller.

A lead-compensation effect (proportional-plus-derivative) can be achieved if the feedback function of Fig. **6-22** is obtained by means of the arrangement of Fig. **6-26.** The feedback function in this case is given by

$$\frac{x_{fb}(s)}{P_o(s)} = K_{fb} \left(\frac{\frac{T_d}{b} s + 1}{T_d s + 1} \right) \tag{6-72}$$

where

$$T_d \triangleq R_d C_d \tag{6-73}$$

$$b \triangleq 1 + \frac{A_d}{A_p} \tag{6-74}$$

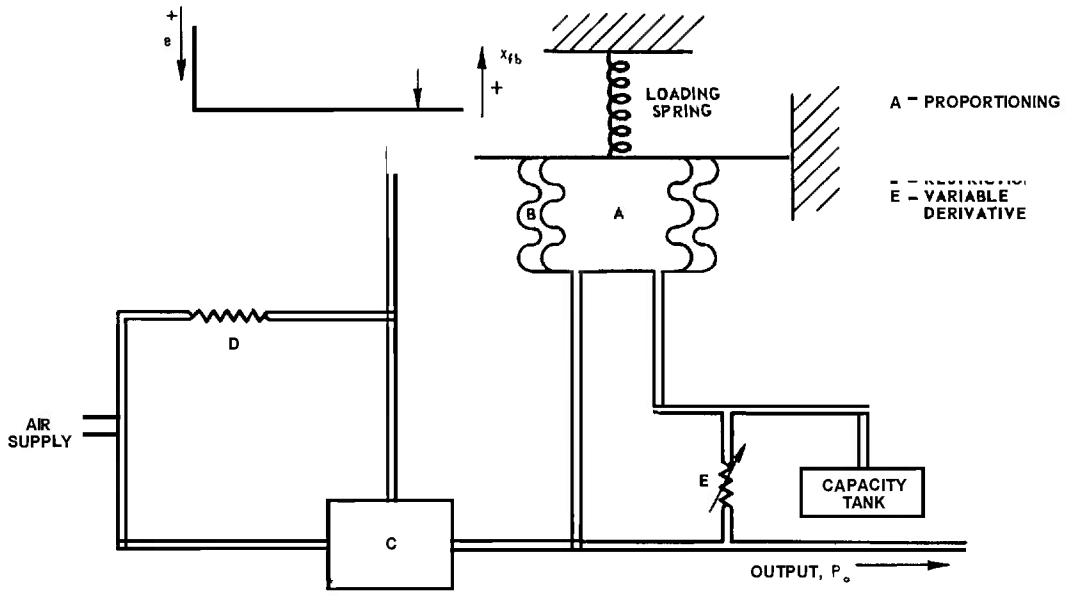
R_d = derivative resistance
C_d = capacitance of tank
A_d = area of derivative bellows
A_p = area of proportional bellows
K_{fb} = sensitivity of proportional bellows

The transfer function of this controller is

$$\frac{P_o(s)}{e(s)} = \frac{(1 - K_1)\, k_{nf} k_{pr}}{1 + K_1 k_{nf} k_{pr} K_{fb}} \left(\frac{T_d s + 1}{\frac{T_d}{\alpha_d} s + 1} \right) \tag{6-75}$$

where

$$\alpha_d = \frac{1 + K_1 k_{nf} k_{pr} K_{fb}}{1 + \frac{1}{b} (K_1 k_{nf} k_{pr} K_{fb})} \tag{6-76}$$



If the product $\frac{1}{b}(K_1 k_{nf} k_{pr} K_{fb})$ is very high (greater than 10), the response of the proportional-plus-derivative controller is approximately

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} \left(\frac{T_d s + 1}{\frac{T_d}{b} s + 1} \right) \qquad (6\text{-}77)$$

If the ratio of the areas (A_d/A_p) is very high (greater than 50), then $b >> 1$, and the response is given approximately by

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} (1 + T_d s) \qquad (6\text{-}78)$$

The form of the right side of this equation explains the term —"proportional-plus-derivative" controller.

BIBLIOGRAPHY

1 Edited by I. A. Greenwood, Jr., J. V. Holdam, Jr., and D. MacRay, Jr., ***Electronic Instruments,*** MIT Radiation Laboratory Series, Vol. **21,** pp. #319-**344,** McGraw-Hill Book Company, Inc., New York, N. Y., **1948.**

2 H. M. James, N. B. Nichols, and R. S. Phillips, ***Theory of Servomechanisms,*** MIT Radiation Laboratory Series, Vol. **25,** pp. #114-130, **196-211,** McGraw-Hill Book Company, Inc., New York, N.Y., **1947.**

3 **G.** S. Brown and D. P. Campbell, *Principles* of ***Servomechanism,*** pp. #195-**235, 262-292,** John Wiley & Sons, Inc., New York, N.Y., **1948.**

4 H. Chestnut and ***R.*** W. Mayer, ***Servomechanisms and Regulating System Design,*** Vol. 1, pp. #245-290, **327-357,** John Wiley & Sons, Inc., New York, N.Y., **1951.**

5 H. Chestnut and R. W. Mayer, ***Servomechanisms and Regulating System Design,*** Vol. *2, pp.* #119-149, **187-213,** John Wiley **&** Sons, Inc., New York, N.Y., **1955.**

6 J. **G.** Truxal, ***Automatic Feedback Control System Synthesis,*** pp. #250-266, **278-409,** McGraw-Hill Book Company, Inc., New York, N.Y., **1955.**

7 W. R. Evans, "Control System Synthesis by Root-Locus Method", ***Trans. AIEE,*** Vol. **69,** Part I, pp. #66-69, **1950.**

8 P. Travers, ***"A*** Note on the Design of Conditionally Stable Feedback Systems", ***Trans, AIEE,*** Vol. 70, Part I, pp. #626-630, **1951.**

9 M. R. Aaron, "Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed Loop Function", ***Trans. AIEE,*** Vol. **70,** Part 11, pp. #1439-1445, **1951.**

10 D. W. Russel and C. H. Weaver, "Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location", ***Trans. AIEE,*** Vol. **71,** Part 11, pp. #95-104, **1952.**

11 J. M. Smith, "Stabilization Templates for Servomechanisms", ***Trans. AIEE,*** Vol. **71,** Part 11, pp. #220-224, **1952.**

12 J. **R.** Koenig, **"A** Relative Damping Criterion for Linear Systems", ***Trans. AIEE,*** Vol. *72,* Part 11, pp. #291-294, **1953.**

13 V. C. **M.** Yeh, "Synthesis of Feedback Control Systems by Gain-Contour and Root-Contour Methods", ***Trans. AIEE,*** Vol. **75,** Part 11, pp. #85-96, **1956.**

14 A. Sobsczyk, "Stabilization of Carrier Frequency Servomechanisms", J. *Franklin Inst.,* Vol. 246, pp. #21-43, 95-121, 187-213, July-August-September, 1948.

15 A. P. Notthoff, Jr., "Phase Lead for A-C Servo Systems with Compensation for Carrier Frequency Changes", *Trans. AIEE,* Vol. 69, Part I, pp. #285-291, 1950.

16 D. McDonald, "Improvements in the Characteristics of A-C Lead Networks for Servomechanisms", *Trans. AIEE,* Vol. 69, Part I, pp. #293-300, 1950.

17 17 R. L. Cosgriff, "Integral Controller for Use in Carrier-Type Servomechanisms", *Trans. AIEE,* Vol. 69, Part 11, pp. #1379-1383, 1950.

18 G. M. Attura, "Effects of Carrier Shifts on Derivative Networks for A-C Servomechanisms", *Trans. AIEE,* Vol. 70, Part II, pp. #612-618, 1951.

19 G. A. Bjornson, "Graphical Synthesis by Graphical Methods for A-C Servomechanisms", *Trans. AIEE,* Vol. 70, Part I, pp. #619-625, 1951.

20 S. S. L. Chang, "Transient Analysis of A-C Servomechanisms", *Trans. AIEE,* Vol. 74, Part 11, pp. #30-37, 1955.

21 L. A. Gould and P. E. Smith, Jr., "Dynamic Behavior of Pneumatic Devices", *Instruments,* Vol. 26, No. 6 and 7, pp. #886-891, 1026-1033, 1953.

CHAPTER 7

PERFORMANCE EVALUATION*

7-1 RELATIONS BETWEEN FREQUENCY RESPONSE AND TRANSIENT RESPONSE

7-1.1 GENERAL

As stated in Par. 6-5, the relationship between transient response and frequency response is somewhat tenuous. Consequently, it is often necessary to have explicit knowledge of the response in both the time and frequency domains. This section presents some of the important approximations that enable the designer to translate between the time- and frequency-domain descriptions of performance. By the use of these approximations, a quick evaluation of performance can be made.

7-1.2 CLOSED-LOOP FREQUENCY RESPONSE FROM CLOSED-LOOP TRANSIENT RESPONSE

If the closed-loop *transient response* of a system is known from experimental test data, there are several methods[21,22,27,28,29,30] available for determining the frequency response.

If the step response of the system is nonoscillatory (i.e., has no overshoot), the transient component of the response can be obtained by subtracting the step response from the final value of the output, i.e.,

$$c_t(t) = c(\infty) - c(t) \tag{7-1}$$

where $c(\infty)$ is the final value of the output, $c(t)$ is the step response, and $c_t(t)$ is the transient component of the response. The logarithm of $c_t(t)$ is plotted against time on semi-log paper. If the response is dominated by an exponential component, the resultant curve plotted on semi-log paper eventually approaches a straight line whose slope corresponds to the magnitude of the dominant time constant. That is, if the dominant transient component is

$$c_{t_1}(t) = Ae^{-t/T}, \tag{7-2}$$

then

$$\log_e c_{t_1}(t) = \log_e A^{-t/T} \tag{7-3}$$

An extrapolation of the straight-line asymptote of $\log_e c_{t_1}(t)$ back to zero time yields the logarithm of the amplitude A of the dominant transient component. Thus, the dominant transient component is completely determined and can be subtracted from $c_t(t)$. A plot of the logarithm of the difference $[c_t(t) - c_{t_1}(t)]$ versus time t produces a curve which approaches a straight-line asymptote whose slope corresponds to the time constant of the exponential component having the next smaller time constant. Extrapolating this curve back to zero time yields the logarithm of the amplitude of the secondary component, $c_{t_2}(t)$. Next, the function $[c_t(t) - c_{t_1}(t) - c_{t_2}(t)]$ is determined, and the process can be repeated until the limit of measurement accuracy is reached.

Thal-Larsen[21] gives a method for determining approximate transfer functions based on the approximation of a nonoscillatory step response by the transfer function

$$W(s) = \frac{C(s)}{R(s)} = \frac{e^{-t_o s}}{(s+1)(T_2 s+1)(T_3 s+1)} \tag{7-4}$$

By L. A. Gould

where T_2 and T_3 are dimensionless time constants and t_0 is the dead time.

By choosing the 10%, 40%, and 80% times in the step responses of this system for the various combinations of its parameters, a set of dimensionless curves have been constructed. In using the curves (Figs. 7-1 through 7-5), the three points corresponding to the 10%, 40%, and 80% response levels of the measured response are determined, and the times corresponding to these points are designated t_1, t_2, and t_3, respectively. The values of the dimensionless ratio $(t_3 - t_1)/(t_2 - t_1)$ and the time $(t_3 - t_1)$ together with curves of Figs. 7-4 and 7-5 will enable the designer to determine a set of roots that corresponds to a transient passing through the three selected points. If a dead time t_0 is present, the ratio $(t_3 - t_1)/(t_2 - t_0)$ will enable the designer to select roots that reproduce the first 10% of the transient.

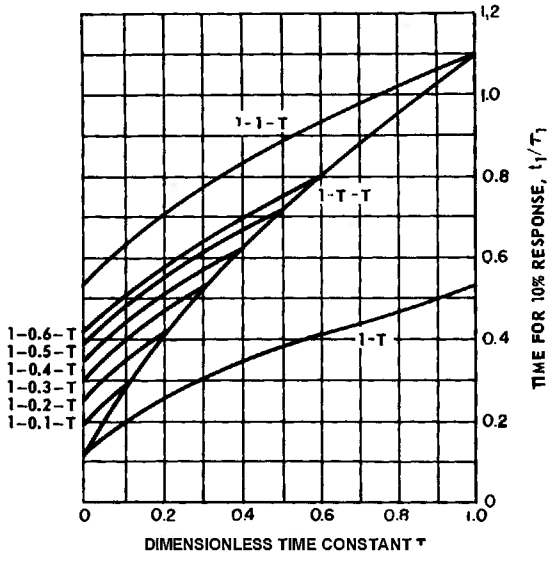


Fig. 7-1 Normalized curves yielding time for 10-percent transient response corresponding to combinations of various time constants

By permission from *Transactions of the AIEE*, Volume 74 Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

Example. Let

$t_1 = 0.97$ sec

$t_2 = 2.14$ sec

$t_3 = 4.47$ sec

Then,

$$\frac{t_3 - t_1}{t_2 - t_1} = 3.00$$

(a) Entering Fig. 7-4 at this value (i.e., 3.00), several curves are crossed allowing the choice of various combinations of the dimensionless or relative time constant T. Choosing three of these combinations :

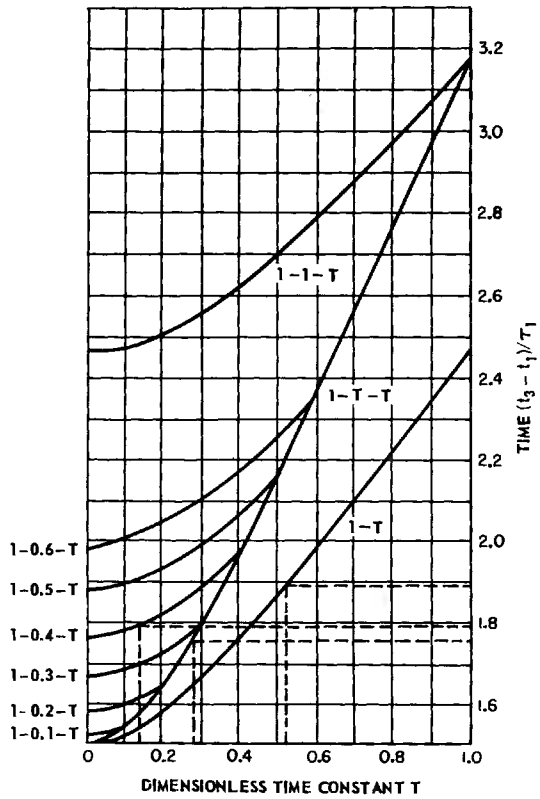


Fig. 7-2 Normalized curves yielding time for 40-percent transient response corresponding to combinations of various time constants.

By permission from *Transactions of the AIEE*, Volume 74, Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

(1) Curve 1—T—T: 1, 0.275, 0.275

(2) Curve 1—**0.4**—T: **1, 0.4,** 0.135

(3) Curve 1—T: 1, 0.520

(b) Entering Fig. 7-2 with the dimensionless time constants found above, the dimensionless time $(t_3 - t_1)/\tau_1$ is determined :

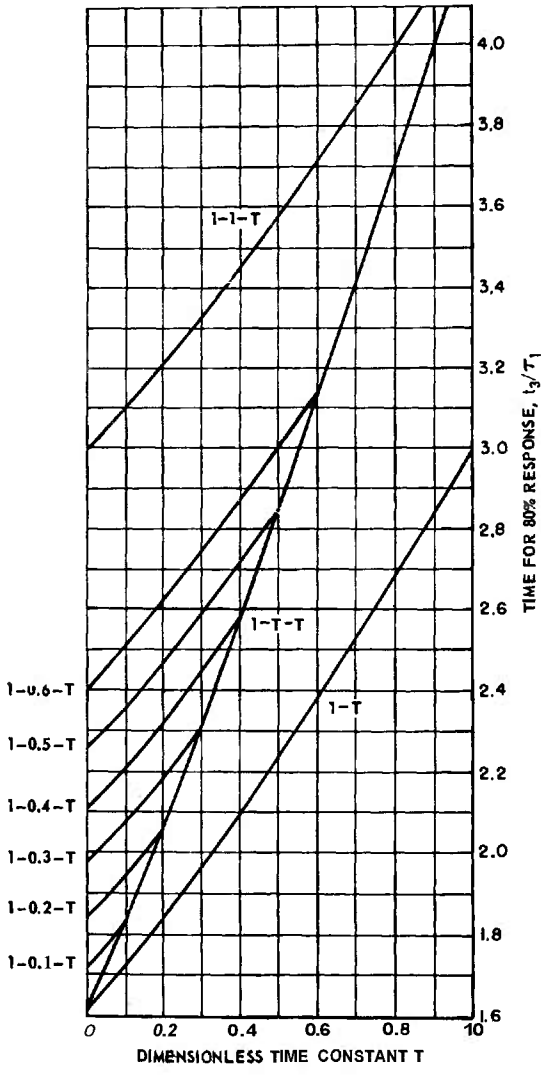


Fig. 7-3 Normalized curves yielding time for 80-percent transient response corresponding to combinations of various time constants.

By permission from *Transactions* of *the AIEE.* Volume 74, Part II, 1966, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

(1) Curve 1—T—T for T = 0.275: 1.755

(2) Curve **1—0.4**—T for T = 0.135: 1.790

(3) Curve 1—T for T = 0.520: 1.890

(c) The time $(t_3 - t_1)$ from the actual transient divided by the dimensionless time $(t_3 - t_1)/\tau_1$ yields the conversion factor τ_1 by which the relative time constants found in the first step must be multiplied to obtain the

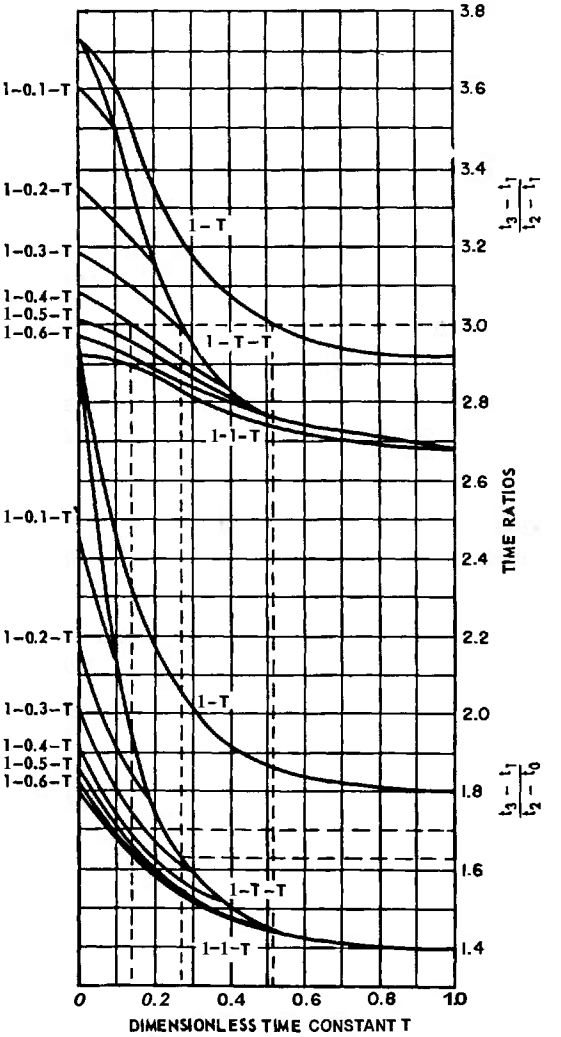


Fig. 7-4 Normalized curves yielding time-interval ratios of the transient response corresponding to combinations of various time constants.

By permission from *Transactions* of *the AIEE.* Volume 74, Part II, 1966, from article entitled. 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

actual time constants. Note that **Eq.** (7-4) represents a normalized transfer function with dimensionless time constants 1, T_2, and T_3. The time constants for the original function before normalization are τ_1, $\tau_1 T_2$, and $\tau_1 T_3$. For the example $(t_3 - t_1) = 3.50$ sec, the three combinations which fit the original curve are:

(1) $\tau_1 = 1.995$ sec; $\tau_1 T_2 = \tau_1 T_3 = 0.549$ sec

(2) $\tau_1 = 1.955$ sec; $\tau_1 T_2 = 0.782$ sec; $\tau_1 T_3 = 0.264$ sec

(3) $\tau_1 = 1.850$ sec; $\tau_1 T_2 = 0.963$ sec; $\tau_1 T_3 = 0$

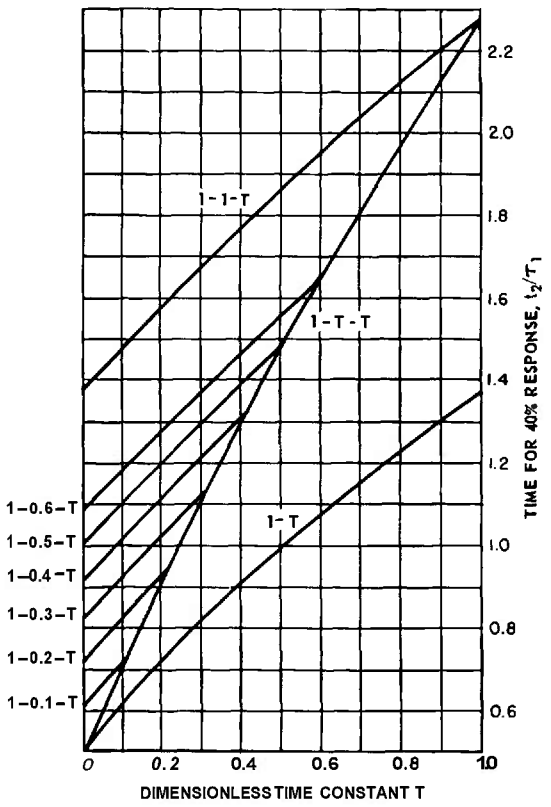


fig. 7-5 Normalized curves yielding the time interval between 70- and 80-percent response of the transient corresponding to combinations of various time constants.

By permission from *Transactions of the AIEE*, Volume 74, Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

(d) To check for the necessity of a dead-time factor, enter Fig. 7-5 with the dimensionless time constants T from the second step to determine t_2/τ_1.

(1) Curve $1 - T - T$ for $T = 0.275$; 1.075

(2) Curve $1 - 0.4 - T$ for $T = 0.135$; 1.055

(3) Curve $1 - T$ for $T = 0.520$; 1.018

(e) The conversion factor τ_1 found in the third step, together with the results of the fourth step, permit the calculation of the actual time t_2 if no dead time is present. Thus, for the three combinations considered, there results

(1) $t_2 = 2.142$ sec

(2) $t_2 = 2.060$ sec

(3) $t_2 = 1.885$ sec

(f) The times found in the fifth step when subtracted from the measured time t_2 yield the dead time t_0. The actual measured time $t_2 = 2.14$ sec. Therefore,

(1) $t_0 = 2.14 - 2.142 = 0$

(2) $t_0 = 2.14 - 2.060 = 0.080$ sec

(3) $t_0 = 2.14 - 1.885 = 0.255$ sec

(g) By substituting the appropriate values from steps *(c)* and (f) into **Eq.** (7-4), we find that the three transfer functions which approximate the response in the 10% to 80% interval are

(1) $W(s) \cong \dfrac{1}{(1.995s+1)\ (0.549s+1)^2}$

(2) $W(s) \cong \dfrac{e^{-0.0808}}{(1.955s+1)\ (0.782s+1)\ (0.264s+1)}$

(3) $W(s) \cong \dfrac{e^{-0.255s}}{(1.850s+1)\ (0.963s+1)}$

Chestnut and Mayer[27] give graphical methods that are useful for determining frequency response from transient response in

any case (oscillatory or nonoscillatory). To find the frequency response associated with the step response of a system, the time axis of the response is divided into equal intervals. Then a "staircase" approximation (see Fig. 7-6) is made to the step response with each step occurring at the middle of a given time interval. If t_n is the middle of the nth time interval and ΔC_n is the change in the response occurring at t_n, then the frequency response is given by

$$W(j\omega) = \sum_{n=1}^{\infty} \Delta C_n e^{-j\omega t_n} \qquad (7\text{-}5)$$

This equation can be evaluated graphically at each frequency by a "vector" summation.

To find the frequency response corresponding to the impulse response of a system (the impulse being approximated experimentally by a short finite pulse), the time scale of the

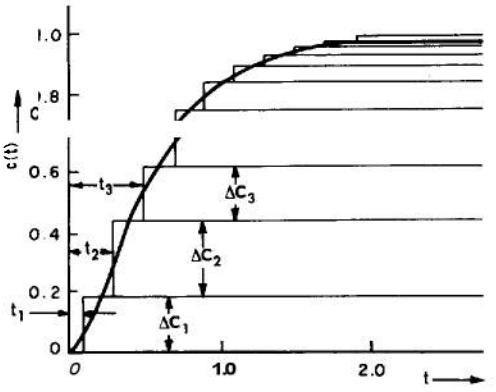


Fig. 7-6 Rectangular approximation to step response.

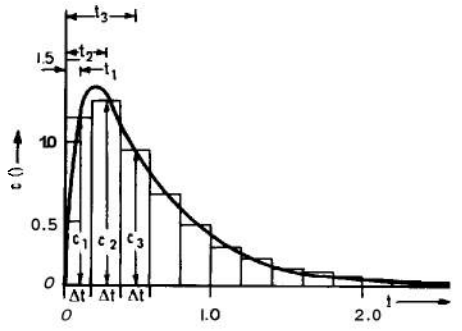


Fig. 7-7 Rectangular approximation to impulse response.

impulse response is divided into equal intervals. Then a rectangular-pulse approximation is made to the impulse response (Fig. 7-7). If t_n is the center of the nth time interval, c_n —the value of the impulse response at t_n, and At — thelength of the time interval, then the frequency response is given by

$$W(j\omega) = \sum_{n=1}^{\infty} c_n \Delta t e^{-j\omega t_n} \qquad (7\text{-}6)$$

This equation can be evaluated graphically at each frequency by a "vector" summation.

Seamans et al.[28,29] use a triangular method to approximate a given time function $c(t)$. This is equivalent to approximating the time function by straight-line segments and then decomposing the straight-line approximation into a series of isosceles triangles (Fig. 7-8). Once the transform of a single triangular pulse is known, the frequency function $C(j\omega)$ corresponding to $c(t)$ is found from

$$C(j\omega) = e^{-j\omega t_0}\left[\frac{\sin\frac{\omega\Delta t}{2}}{\frac{\omega\Delta t}{2}}\right]^2 \Delta t\left[\sum_{n=1}^{\infty} E_n e^{-jn\omega\Delta t}\right] \qquad (7\text{-}7)$$

where t_0 represents the time at the start of the first pulse, At — the time interval between pulses, and E_n —the amplitude of the nth pulse.

Guillemin[28,30] suggests that the time function be approximated by a sequence of rational polynomials in t (straight lines, parabolas,

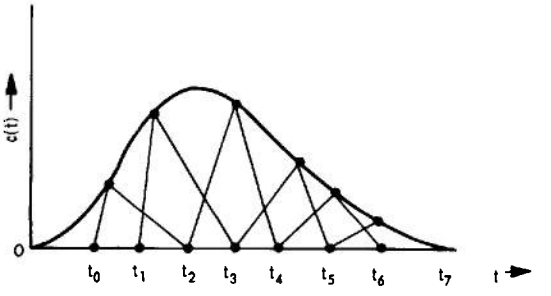


Fig. 7-8 Triangular approximation to time function.

cubics, etc.). The approximate function $c^*(t)$ is differentiated enough times (n times) to make $\frac{d^n}{dt^n}c^*$ a sequence of impulses. Actually, the original function may be differentiated before approximating by polynomials so that lower-order polynomials can be used. The final impulse function is then transformed, yielding

$$C(j\omega) \cong \sum_{k=1}^{m} \frac{a_k e^{-j\omega t_k}}{(j\omega)^n} \qquad (7\text{-}8)$$

where a_k is the magnitude of the kth impulse, t_k — the time of occurrence of the kth impulse, and n — the *total* number of times the original function has been differentiated.

If *rational approximations* are sought for an experimentally derived frequency function $W(j\omega)$, advantage can be taken of the fact that the plot of 10 $\log_{10} |W(j\omega)|$ vs $\log\omega$ is easily representable by straight-line asymptotes having slopes of $\pm 10n$ dg/dec ($n = 0, 1, 2, \ldots$). By combining the straight-line approximation of the magnitude function with the first- and second-order response curves given in Par. 5-3 (Figs. 5-7 through 5-10), curve fitting is possible. The easiest procedure is to use the magnitude curves to get a rough approximation and then to refine the approximation with the phase curves.

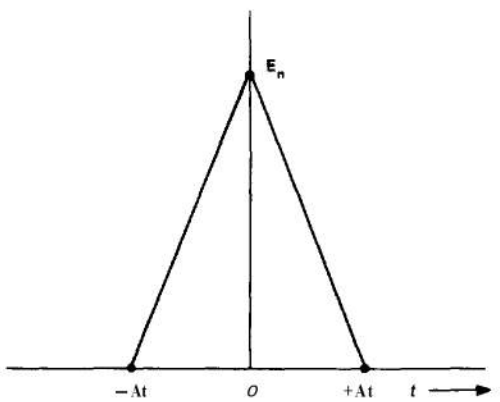


Fig. 7-9 Triangle function.

Linvill[15,28] has proposed a method for improving the foregoing approximation procedure. In this method, an investigation is made of the effect of varying the position of the approximate poles and zeros on the difference between the actual function and the first approximation obtained from fitting the asymptotes and their corresponding correction curves. For example, if

$$F(\omega) = 10 \log_{10} |G(j\omega)| \qquad (7\text{-}9)$$

and

$$G(s) = \frac{s^2 - 2\sigma_1 s + \sigma_1^2 + \omega_1^2}{s^2 - 2\sigma_2 s + \sigma_2^2 + \omega_2^2} \qquad (7\text{-}10)$$

then the change in $F(\omega)$ resulting from small changes of the poles $(+\sigma_2 \pm j\omega_2)$ and the zeros $(+\sigma_1 \pm j\omega_1)$ is given by

$$\Delta F(\omega) = \frac{\partial F}{\partial \sigma_1}\Delta\sigma_1 + \frac{\partial F}{\partial \omega_1}\Delta\omega_1 + \frac{\partial F}{\partial \sigma_2}\Delta\sigma_2 + \frac{\partial F}{\partial \omega_2}\Delta\omega_2 \qquad (7\text{-}11)$$

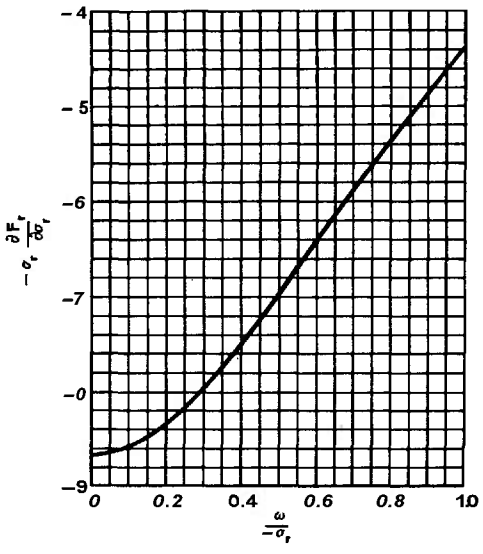
The steps in the approximation procedure are as follows:

(a) A plot is made of the difference between the actual $F(\omega)$ and the first approximation in a frequency region where the first approximation is to be improved by changing the position of approximate poles and/or zeros which occur in this region.

(b) The variation of the pertinent partial derivatives of $F(\omega)$ with frequency ω is determined in the vicinity of the approximate poles and/or zeros.

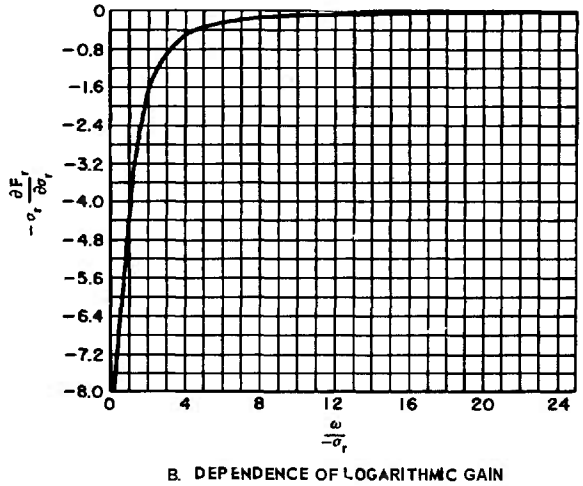
(c) The pertinent partial-derivative curves are used to approximate $\Delta F(\omega)$ in the frequency region of interest. From this approximation, the necessary changes in the positions of the approximate poles and/or zeros are determined.

The curves of Figs. 7-10A through 7-10N can be used to evaluate the necessary partial-derivative curves. Note that the phase corrections can be determined by using the same procedure.



A. DEPENDENCE OF LOGARITHMIC GAIN ON MOTION OF A REAL ZERO

$F_r = 20 \log |j\omega - \sigma_r|$

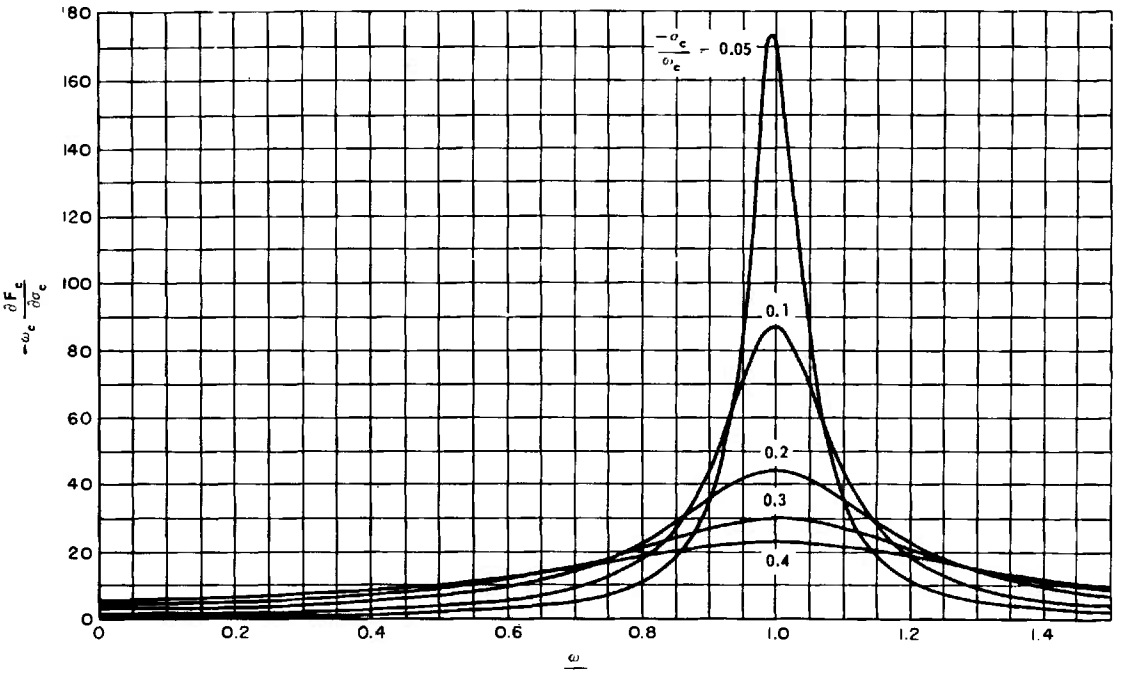


B. DEPENDENCE OF LOGARITHMIC GAIN ON MOTION OF A REAL ZERO

$F_r = 20 \log |j\omega - \sigma_r|$

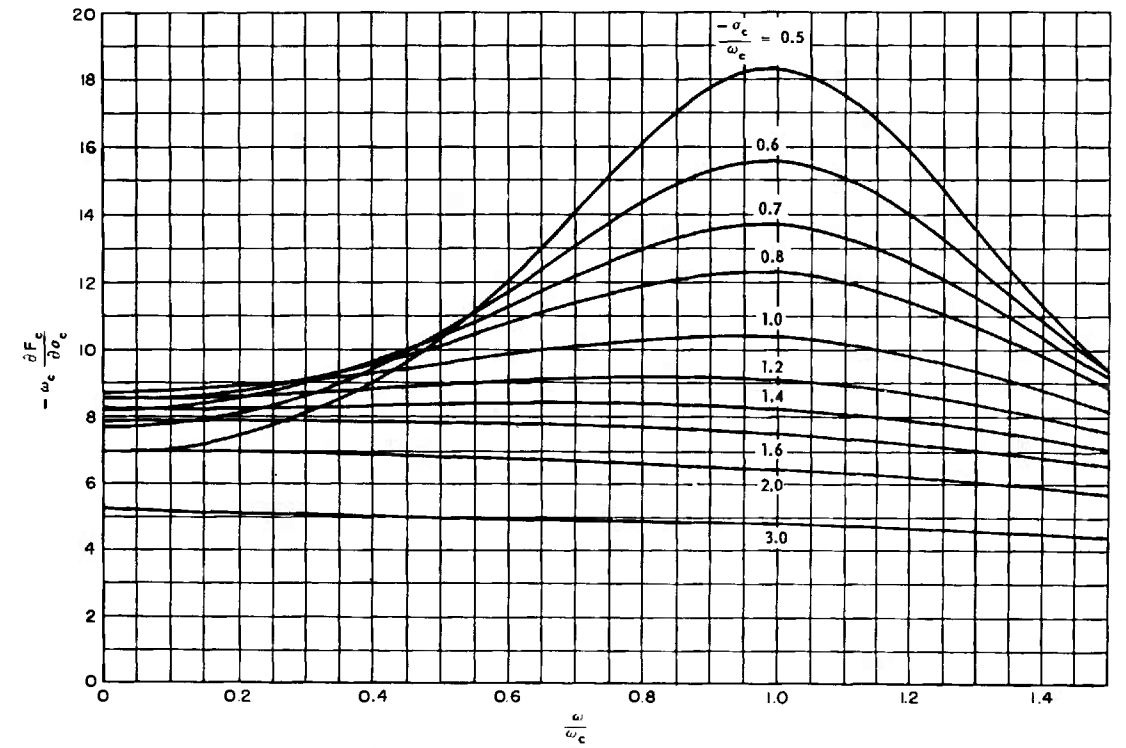
Fig. 7-70 Partial derivatives for Linvill's procedure. (Sheet 7 of 10)

From "The Selection of Network Functions to Approximate Prescribed Frequency Characteristics", by J. G. Linvill, Research Laboratory of Electronics, Technical Report No. 146, March 14, 1960, Massachusetts Institute of Technology.



C. DEPENDENCE OF GAIN ON REAL PART OF ZERO (OR POLE)

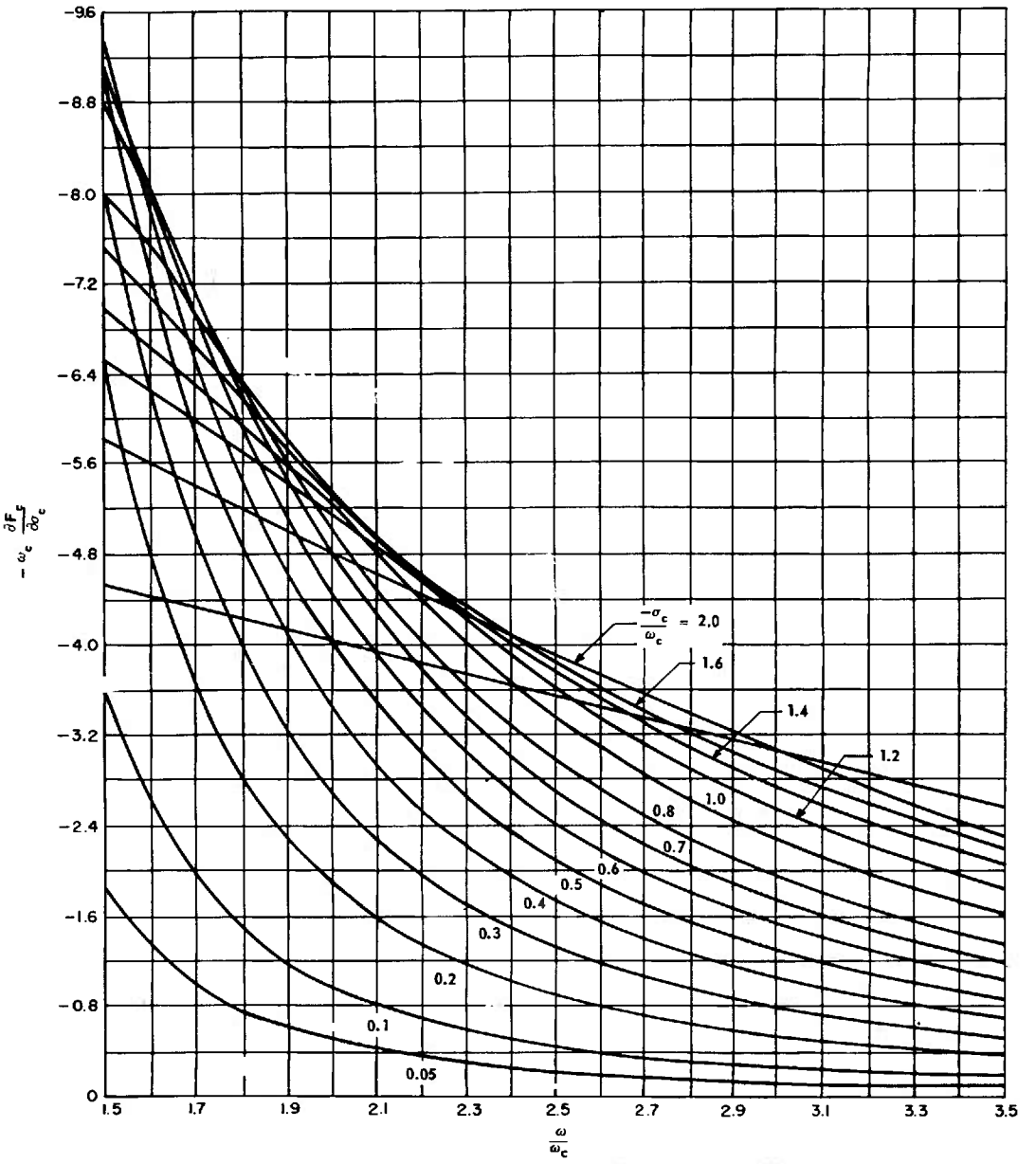
$F_c = 20 \text{ LOG } |(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2|$



D. DEPENDENCE OF GAIN ON REAL PART OF ZERO (OR POLE)

$F = 20 \text{ LOG } |(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2|$

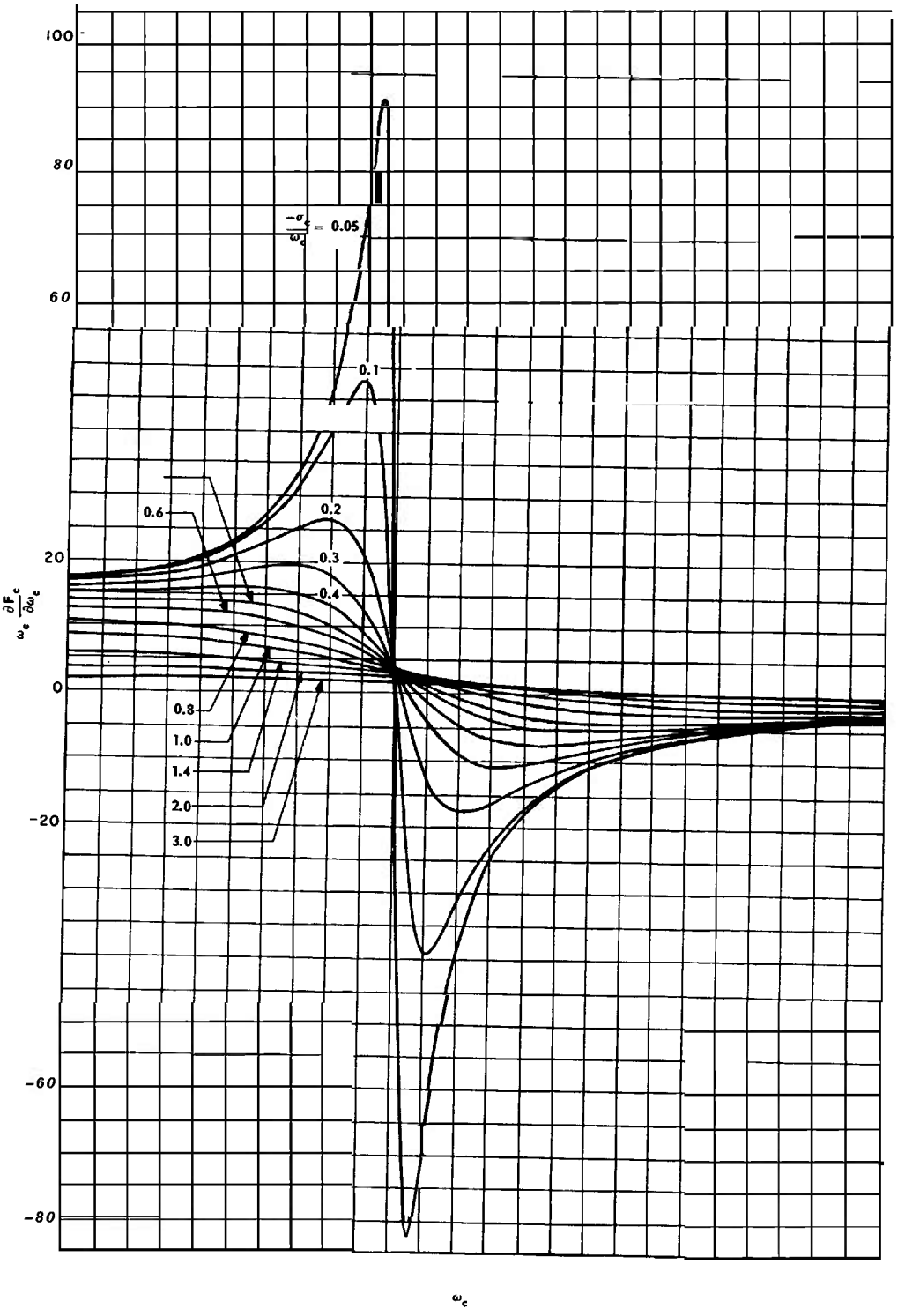
Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 2 of 10)

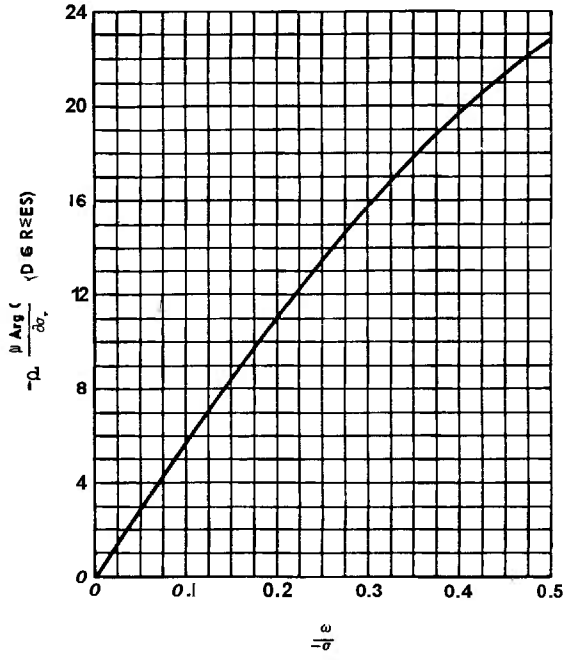


E. DEPENDENCE OF GAIN ON REAL PART OF ZERO (OR POLE)

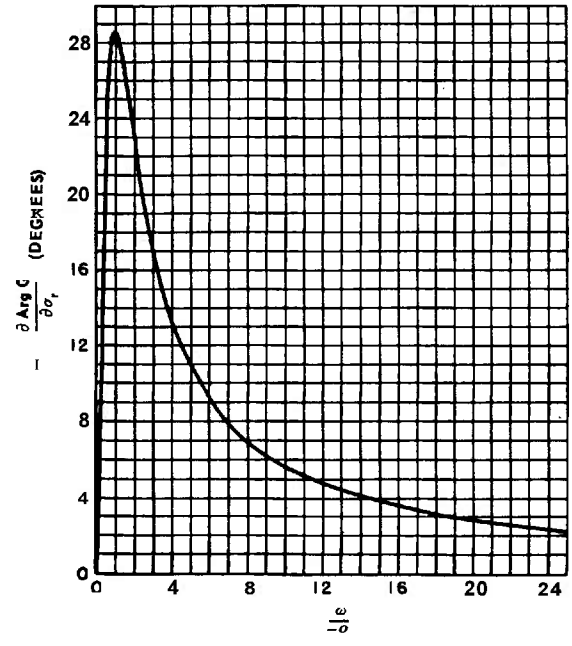
$F_c = 20 \text{ LOG } [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$

Fig. 7-70 Partial derivatives for Linvill's procedure. (Sheet 3 of 70)


100
80
60
20
0
-20
-60
-80
$\frac{-\sigma}{\omega_c} = 0.05$
0.1
0.2
0.3
0.4
0.6
0.8
1.0
1.4
2.0
3.0
$\frac{\partial F_c}{\partial \omega_c} \omega_c$
ω_c

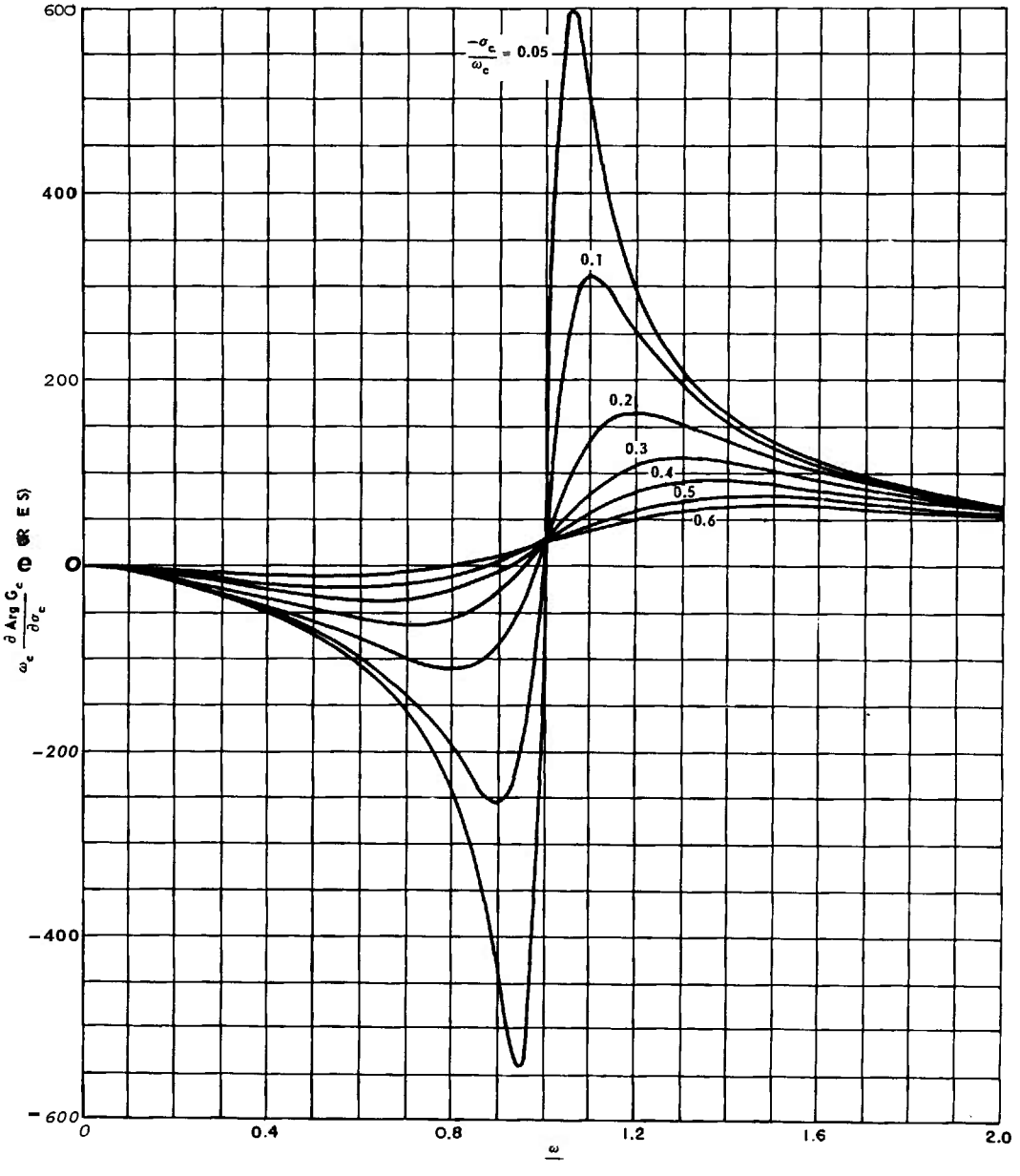


G. DEPENDENCE OF PHASE ON REAL ZERO (OR POLE)
$G_r = (j\omega - \sigma_r)$



H. DEPENDENCE OF PHASE ON REAL ZERO (OR POLE)
$G_r = (j\omega - \sigma_r)$

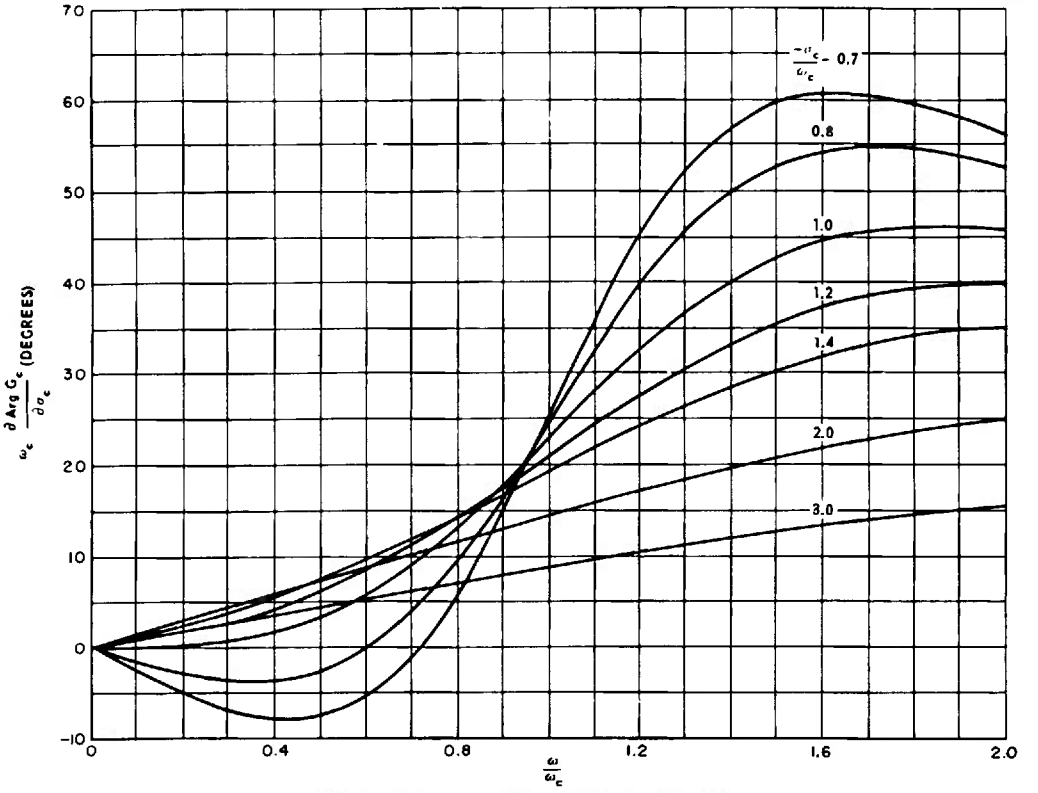
Fig. 7-70 Partial derivatives for linvill's procedure. (Sheet 5 of 70)



I. DEPENDENCE OF PHASE ON REAL PART OF ZERO (OR POLE)

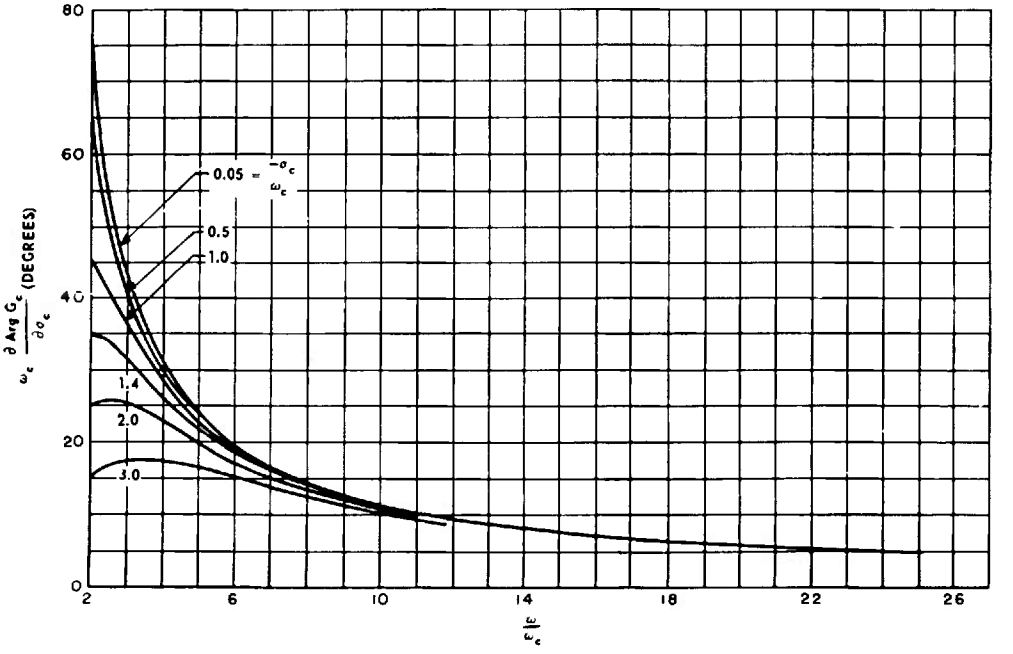
$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$

Fig. 7-70 Partial derivatives for Linvill's procedure. (Sheet 6 of 70)



J. DEPENDENCE OF PHASE ON REAL PART OF ZERO (OR POLE)

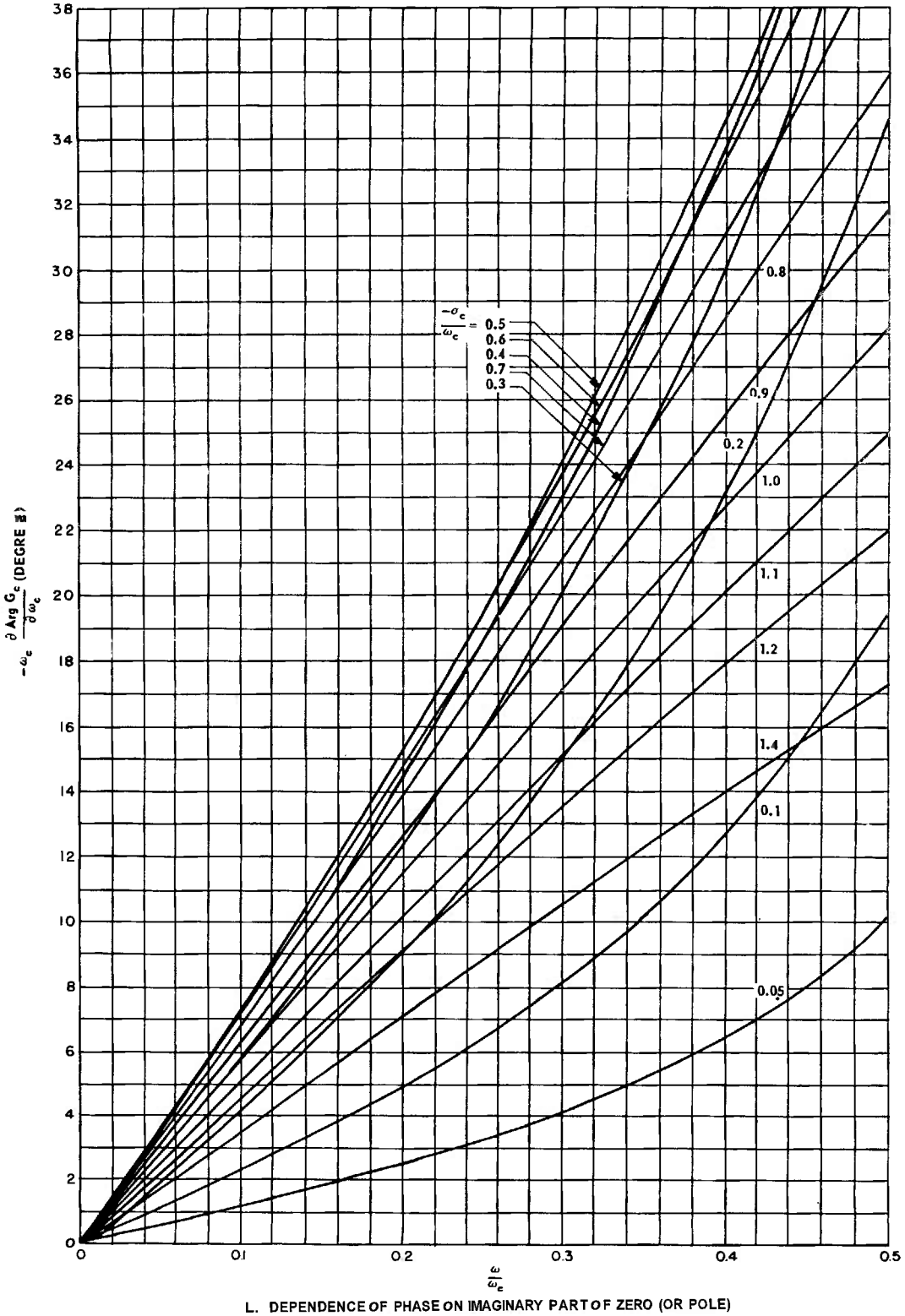
$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$



K. DEPENDENCE OF PHASE ON REAL PART OF ZERO (OR POLE)

$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$

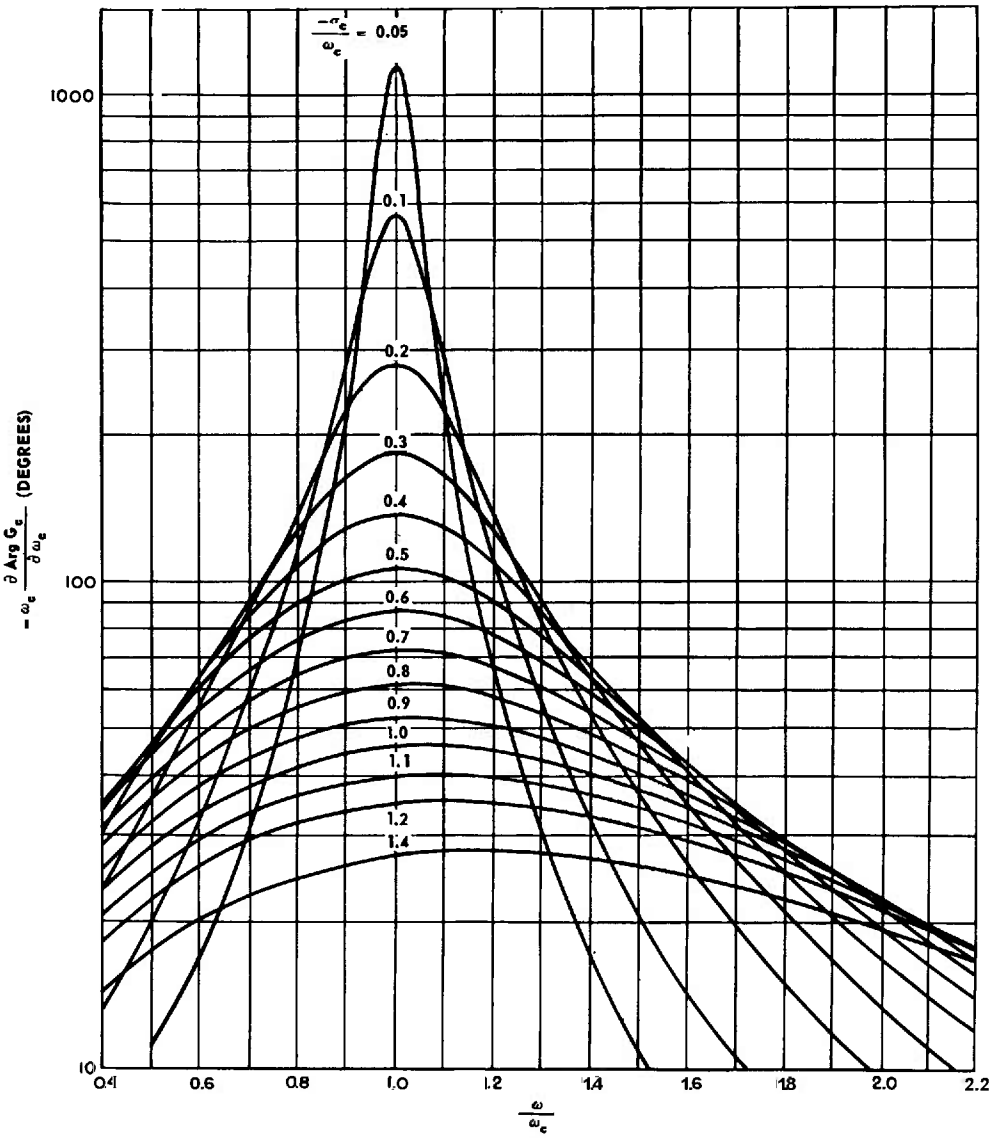
Fig. 7-10 Partial derivatives for *Linvill's procedure. (Sheet 7* of *10)*



L. DEPENDENCE OF PHASE ON IMAGINARY PART OF ZERO (OR POLE)

$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$

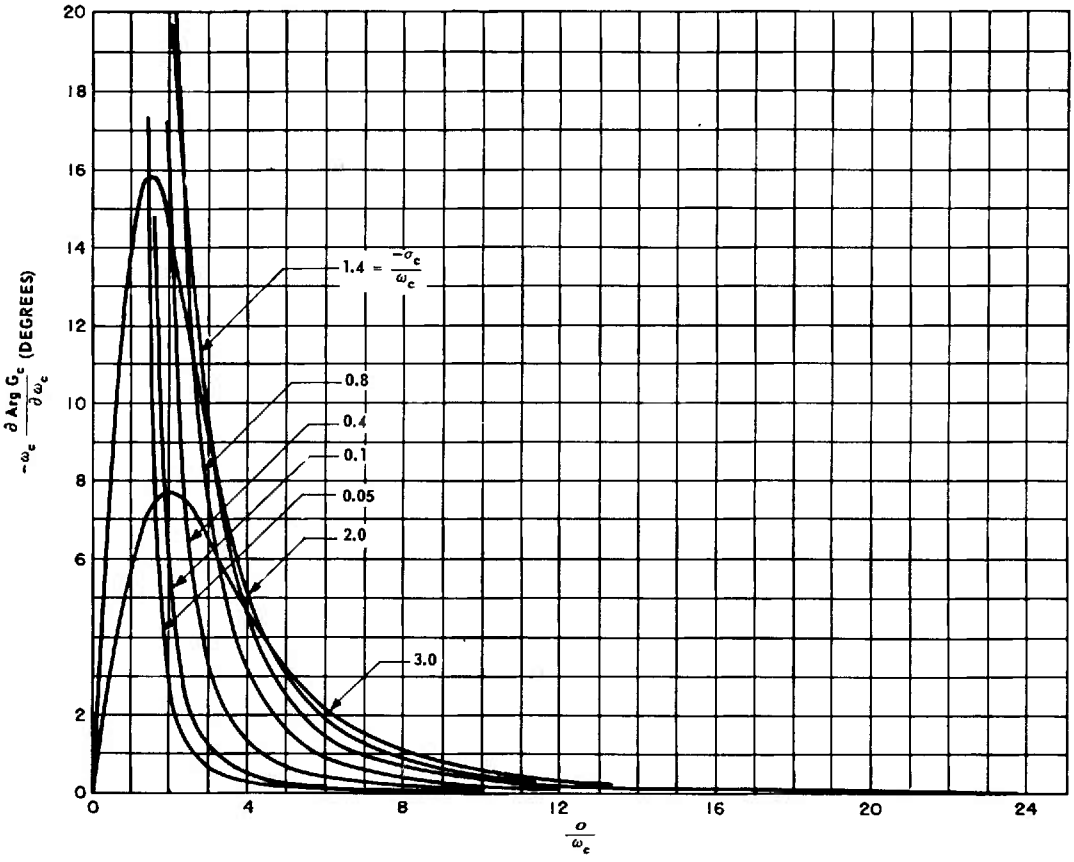
Fig. 7-70 Partial derivatives for linvill's procedure. (Sheet 8 of 70)



M DEPENDENCE OF PHASE ON IMAGINARY PART OF ZERO (OR POLE)

$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$

Fig. 7-70 Partial derivatives for Linvill's procedure. (Sheet 9 of 10)



N. DEPENDENCEOF PHASE ON IMAGINARY PART OF ZERO (OR POLE)

$G_i = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$

Fig. 7-70 Partial derivatives for Linvill's procedure. (Sheet *70* of *10*)

7-1.3 RELATIONS BETWEEN CLOSED-LOOP TRANSIENT RESPONSE AND CLOSED-LOOP POLE-ZERO CONFIGURATION

It is desirable to be able to describe properties of the transient response of a system when one is given the closed-loop pole-zero configuration and vice versa. Usually, the designer is given specifications for some form of the transient response of a system. As a result, the conversion of the transient-response specifications to a desired closed-loop pole-zero configuration is a starting point in many design procedures [see Par. **6-5** and references[3,5,6,19,20,28]]. Since the usual assumption in these design procedures is that the closed-loop performance of the system is primarily controlled by a dominant pair of complex poles (dominant quadratic factor in the denominator), only the characteristics of an underdamped second-order system are presented here.

If the system being examined is a unity-feedback system with a pair of complex-conjugate poles and no closed-loop zeros, the closed-loop transfer function relating output to input is

$$W(s) = \frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (7\text{-}12)$$

The error-to-input transfer function is

$$\frac{E(s)}{R(s)} = \frac{s(s + 2\zeta\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (7\text{-}13)$$

The open-loop transfer function is

$$G(s) = \frac{C(s)}{E(s)} = \frac{\omega_n^2}{s(s + 2\zeta\omega_n)} \quad (7\text{-}14)$$

In these equations,

ω_n = natural frequency

and

ζ = damping ratio.

The magnitude and phase of the closed-loop frequency response $W(j\omega)$ are the second-order quadratic factor curves presented in Par. **5-3.** The velocity constant K_v of the system is

$$K_v = \frac{\omega_n}{2\zeta} \quad (7\text{-}15)$$

The first three error coefficients are

$$e_0 = 0 \quad (7\text{-}16)$$

$$e_1 = \frac{1}{K_v} \quad (7\text{-}17)$$

$$e_2 = \frac{1 - 452}{\omega_n^2} \quad (7\text{-}18)$$

The error response curves for a *unit-ramp* input are given in Fig. **7-11.** Note the steady-state error for a *unit-ramp* input to this system is given by

$$e_{ss} = \frac{2\zeta}{\omega_n} = \frac{1}{K_v} \quad (7\text{-}19)$$

The error response curves for a *unit-step* input are given in Fig. **7-12.** The output response can be obtained from these curves by subtracting them from unity. *The solution time* **or** *settling time t,* of the step response is the time for the output to reach **98%** of its final value or for the error to fall to **2%** of its initial value. For the second-order system,

$$t_s = \frac{4}{\zeta\omega_n} \quad (7\text{-}20)$$

The output response curves for a *unit-step* input are plotted in Fig. **7-13.**

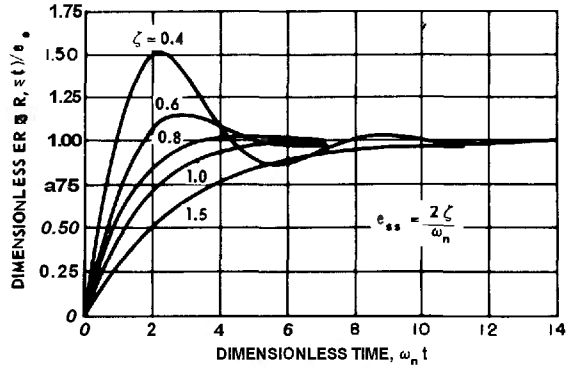


Fig. 7-11 Dimensionless transient error-response curves of a second-order servomechanism to a unit-ramp input.

Adapted with permission from *Principles of Servomechanisms,* by *G. S.* Brown and D. P. Campbell, Copyright. 1948, John Wiley & Sons, Inc.

Quantitative descriptions of the relationships between properties of the transient response and the frequency response of a second-order system will now be given. The resonant frequency of the closed-loop response $W(j\omega)$ is

$$\omega_R = \omega_n \sqrt{1 - 2\zeta^2} \qquad (7\text{-}21)$$

The magnitude of the resonant peak M_p (see Fig. 7-14) is given by the relation

$$M_p = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \qquad (7\text{-}22)$$

The frequency of damped transient oscillation ω_d (damped natural frequency) is

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} \qquad (7\text{-}23)$$

The time taken to reach the first peak in the output response to a unit-step input is

$$t_{m_1} = \frac{\pi}{\omega_n \sqrt{1-\zeta^2}} \qquad (7\text{-}24)$$

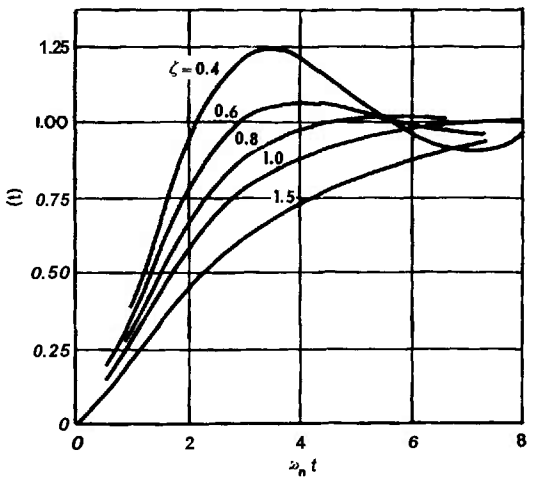


Fig. 7-13 Transient output-response curves of a second-order servomechanism to a unit-step input.

Adapted with permission from *Principles of Servomechanism*, by G. S. Brown and D. P. Campbell. Copyright, 1948. John Wiley & Sone, Inc.

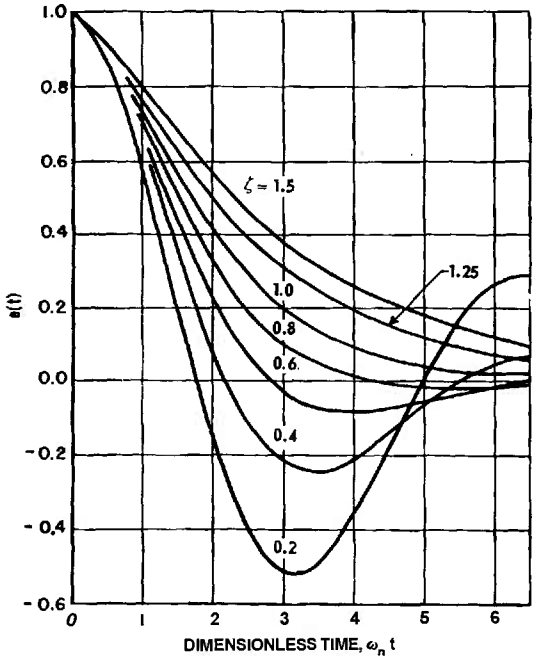


Fig. 7-72 Transient error-response curves of a second-order servomechanism to a unit-step input.

Adapted with permission from *Principles of Servomechanism*, by G. S. Brown and D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

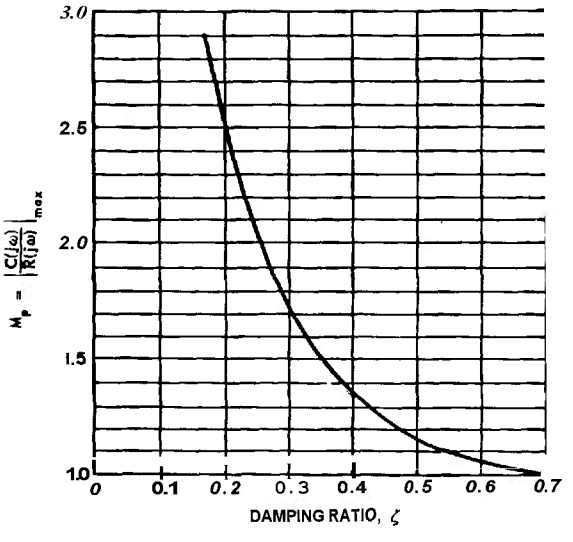


Fig. 7-74 Peak magnifude M_p versus damping ratio ζ for a second-order servomechanism.

Adapted with permission from *Principles of Servomechanisms*, by G. S. Brown and D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

The peak overshoot P_{ov} in the output response to a unit-step input (see Fig. 7-15) is

$$P_{ov} = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}} \tag{7-25}$$

If bandwidth is defined as the frequency ω_b at which $10 \log_{10} |W(j\omega)|$ is down 1.5dg from the zero-frequency value, then

$$\omega_b = \omega_n\left[1 - 2\zeta^2 + \sqrt{2 - 4\zeta^2 + 4\zeta^4}\right]^{1/2} \tag{7-26}$$

With the important characteristics of a second-order system described, it is possible to use these characteristics to aid in establishing a desired closed-loop pole-zero configuration from the transient-response specifications.

A few ***general*** relations between pole-zero configurations and transient-response characteristics are in order. Most closed-loop response functions of unity-feedback systems, $W(s)$, are characterized by a pair of dominant complex poles, one or more dipoles (pole and zero close together), one or more finite zeros, and poles whose magnitudes are much greater (a factor of five or more) than the magnitude of the dominant pair of poles.[3,5,6,12,18,19,28] The pertinent relations are as follows:

(a) The addition of a real zero to $W(s)$ tends to increase the overshoot of the output response to a unit-step input, decreasing the rise time and time delay.

(b) The addition of a real pole to $W(s)$ tends to decrease the overshoot of the output response to a unit-step input, increasing the rise time and time delay.

(c) The addition of real poles to $W(s)$ whose magnitudes are much larger than the magnitude of the dominant pole pair has very little effect on the transient response.

(d) The addition of complex poles to $W(s)$ whose magnitudes are much larger than the magnitude of the dominant pole pair has very little effect on the transient response provided the damping ratio of the added poles is not too small.

(e) The addition of a dipole to $W(s)$ has very little effect on the step response of the system but may have a pronounced effect on the steady-state errors **of** the system.

(f) The excess of poles over zeros for $W(s)$ is equal to or greater than the excess of poles over zeros for the fixed-element transfer function $G_f(s)$.

(g) Most military applications require that $W(0) = 1$. This implies that the open-loop transfer functions $C(s)/E(s) = G(s)$ have at least one pole at the origin.

(h) In any system with one open-loop pole at the origin, the first three error coefficients are

$$e_0 = 0 \tag{7-27}$$

$$e_1 = \frac{1}{K_v} = \sum_{i=1}^{n} \frac{1}{p_i} - \sum_{i=1}^{m} \frac{1}{z_i} \tag{7-28}$$

$$e_2 = -\frac{1}{2}\left(\frac{1}{K_v^2} + \sum_{i=1}^{n} \frac{1}{p_i^2} - \sum_{i=1}^{m} \frac{1}{z_i^2}\right) \tag{7-29}$$

where $-p_j$ is the jth pole of $W(s)$, $-z_j$ is the jth zero of $W(s)$, and K_v is the velocity constant of the system.

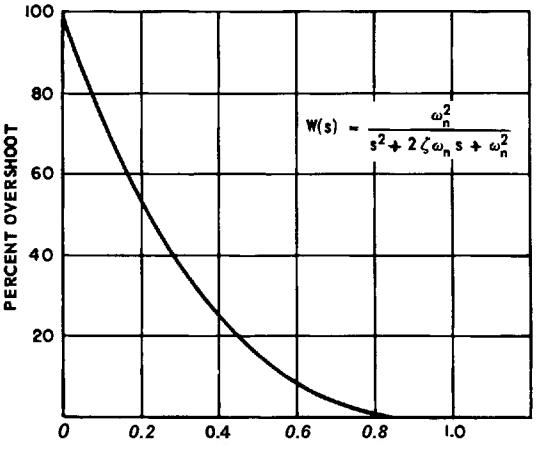


Fig. 7-15 Overshoot variation with ζ.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal. Copyright, 1966, McGraw-Hill Book Company, Inc.

(i) In any system with two open-loop poles at the origin, the first three error coefficients are

$$e_0 = 0 \tag{7-30}$$

$$e_1 = 0 \ (K_v \to \infty) \tag{7-31}$$

$$e_2 = -\frac{1}{2}\left(\sum_{j=1}^{n}\frac{1}{p_j^2} - \sum_{j=1}^{m}\frac{1}{z_j^2}\right) = \frac{1}{K_a} \tag{7-32}$$

where K_a is the acceleration constant, $-p_j$ is the jth pole of $W(s)$, and $-z_j$ is the jth zero of $W(s)$.

(j) If the cutoff frequency ω_{co} is defined as the frequency at which the phase of the closed-loop frequency response is $-90°$ and the buildup time t_{bu} is the time for the output response first to cross unity for a step input, then

$$t_{bu} \cong \frac{\pi}{\omega_{co}} \tag{7-33}$$

(k) If the rise time t_r is defined as the time required for the output response to a unit-step input to go from 10% to 90% of its final value and ω_b is the bandwidth as defined immediately above Eq. (7-26), then for a response with less than 10% overshoot,

$$\frac{\omega_b t_r}{2\pi} \cong 0.30 \text{ to } 0.45 \tag{7-34}$$

(l) If the delay time t_d is defined as the time for the output response to a unit step to reach 50% of its final value, then

$$t_d \cong \frac{1}{K_v} \tag{7-35}$$

(m) If the rise time t_r is as defined in (k),

$$\frac{1}{2\mathrm{x}} t_r^2 \cong -\left(2e_2 + \frac{1}{K_v^2}\right) \tag{7-36}$$

where e_2 is the second error coefficient and K_v is the velocity constant.

(n) If the settling time (solution time) t_s is the time for the output response to a unit-step input to reach 98% of its final value, then

$$t_s \cong 3t_{bu} \text{ to } 5t_{bu} \tag{7-37}$$

From the characteristics of the second-order system and the general relations (a) through (n) of the preceding paragraph, the conversion of time-domain specifications to a closed-loop pole-zero configuration becomes a fairly straightforward matter. Truxal[28] presents a very good description of a typical procedure.

Example. The specifications for a servomechanism are as follows:

(a) $G_f(s) = \dfrac{K}{s(s+a)}$

(b) The bandwidth ω_b of the closed-loop response shall be less than 50 rad/sec, and the output response of the system to a unit-step input shall have an overshoot less than 5% of the final value.

(c) $K_v \geqq 50 \text{ sec}^{-1}$

(d) $|e_2| \leqq 0.01 \text{ sec}^2$

Find a closed-loop pole-zero configuration that satisfies these specifications.

Solution. If the system is initially approximated by a second-order response with no zeros, Fig. 7-15 shows that $\zeta \geqq 0.7$ for a peak overshoot $P_{ov} \leqq 5\%$. For $\zeta = 0.7$, Eq. (7-26) yields $\omega_b = \omega_n$. Therefore, $\omega_n \leqq 50$ rad/sec. The dominant pole pair is thus placed at $s = -35 \pm 35\mathrm{j}$ corresponding to $\omega_n = 49.5$ rad/sec and $\zeta = 0.707$. Now using Eq. (7-15), we find that $K_v = 35 \text{ sec}^{-1}$, which is too small a value. To increase K_v, a dipole will be added. The pole of the dipole must not be set at too low a frequency or else an excessively long tail in the transient will occur. The magnitude of the pole of the dipole will therefore be chosen to be one-tenth the real part of the dominant poles. This corresponds to a pole at -3.5. To determine the location of the zero of the dipole, Eq. (7-28) is used. At this point, the approximate closed-loop response is

$$W(s) \cong \frac{\dfrac{p_1}{z_1}\omega_n^2(s+z_1)}{(s^2+2\zeta\omega_n s+\omega_n^2)(s+p_1)}$$

where $p_1 = +3.5$, $\zeta = 0.7$, $\omega_n = 49.5$, and z_1 is to be determined. Using Eq. (7-28),

$$\frac{1}{K_v} - \frac{1}{p_1} + \frac{1}{\zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}} + \frac{1}{\zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}} - \frac{1}{z_1}$$

or,

$$\frac{1}{z_1} = \frac{1}{p_1} + \frac{2\zeta}{\omega_n} - \frac{1}{K_v}$$

Therefore, $z_1 = 3.40$. The desired pole-zero configuration for $W(s)$ is given by

$$W(s) \cong \frac{2520\,(s+3.4)}{(s^2+70s+2450)\,(s+3.5)}$$

The second error coefficient of this system can be found from Eq. (7-29). Thus,

$$e_2 = -\frac{1}{2}\left[\frac{1}{K_v^2} + \frac{1}{p_1^2} + \left(\frac{1}{\zeta\omega_n + j\sqrt{1-\zeta^2}}\right)^2 + \left(\frac{1}{\zeta\omega_n - j\sqrt{1-\zeta^2}}\right)^2 - \frac{1}{z_1^2}\right]$$

Evaluating this expression, it is found that $e_2 = 2 \times 10^{-3}$ sec^2, which is well within specifications. All the specifications should be checked at this point to insure that the system behaves as desired. The example above has been carried out far enough to demonstrate the basic ideas involved in finding a closed-loop pole-zero configuration that satisfies the given specifications.

7-1.4 RELATIONS BETWEEN OPEN-LOOP FREQUENCY RESPONSE AND CLOSED-TRANSIENT RESPONSE[4,7,8,11,12,25,26]

Since most of the design techniques discussed in Pars. 6-2, 6-3, and 6-4 involve considerations of the open-loop frequency response $C(j\omega)/E(j\omega) = G(j\omega)$, methods for relating the open-loop frequency response to the closed-loop transient response will be presented here.

Harris et al.[7] present an approximate technique for determining the error response $e(t)$ to a transformable input $r(t)$. If ω_c is defined as the frequency at which the open-loop asymptotes cross 0 dg (asymptote crossover frequency ; see Fig. 7-16), this method assumes that ω_c occurs in a region where the slope of the asymptote is −10 dg/dec. In general, the shape of the open-loop asymptote for frequencies greater than ω_c has little effect on the transient response of the system.

The reciprocal error-to-input transfer function $R(s)/E(s)$ can be found from the open-loop response $C(s)/E(s)$ by using the relation

$$\frac{R(s)}{E(s)} = \frac{C(s)}{E(s)} + 1 \qquad (7\text{-}38)$$

Since the open-loop asymptote function $C(s)/E(s)$ is almost always a monotonically decreasing function of frequency, the asymptote crossover frequency ω_c divides the frequency scale into two regions :

$$\frac{R(s)}{E(s)} \cong \frac{C(s)}{E(s)} \text{ for } \omega << \omega_c \qquad (7\text{-}39)$$

$$\frac{R(s)}{E(s)} \cong 1 \text{ for } \omega >> \omega_c \qquad (7\text{-}40)$$

The procedure for finding $e(t)$ is as follows :

(a) From $[R(s)/E(s)]_{approximate}$ by using all factors of $C(s)/E(s)$ corresponding to poles and zeros of $C(s)/E(s)$ with magnitudes (break frequencies) less than ω_c. Delete all other factors of $C(s)/E(s)$ and add a numerator factor equal to $\left(1+\frac{s}{\omega_c}\right)$.

(b) $[E(s)/R(s)]_{approximate}$ is the reciprocal of $[R(s)/E(s)]_{approximate}$. From the transform $R(s)$ of the input $r(t)$ and the approximate error-to-input transfer function, find the first

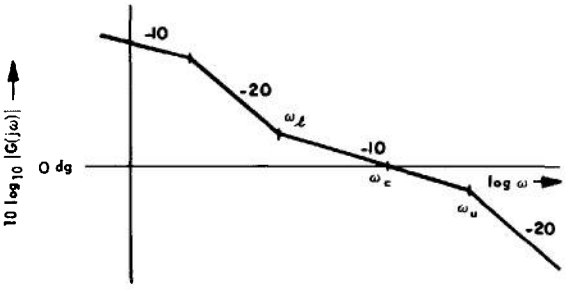


Fig. 7-76 Typical open-loop asymptote function.

approximation to $e(t)$ by performing the inverse Laplace transformation of the function $[E(s)/R(s)]_{approximate} \times R(s)$ and plotting this time function.

(c) Find the correction ratio,

$$\rho = \left(\frac{[R(s)/E(s)]_{approximate}}{[R(s)/E(s)]_{exact}} \right)_{s = j\omega_c} \qquad (7\text{-}41)$$

(d) The ends ω_l and ω_u of the -10 dg/dec slope region which fixes ω_c are called the lower-and upper-corner (or break) frequencies of the -10 dg/dec region (see Fig. 7-16). The plot of the first approximation to $e(t)$ found in (b) is multiplied by the correction ratio ρ in the time interval

$$\frac{1}{\omega_u} < t < \frac{1}{\omega_l} \qquad (7\text{-}42)$$

and the resulting curves are joined smoothly in the regions $t = 1/\omega_u$ and $t = 1/\omega_l$. This method works best if the -10 dg/dec slope region is fairly long ($\omega_u/\omega_l \cong 8$) and if the closed-loop M_p is close to unity.

Chestnut and Mayer [8,26] present a series of charts that can be used to determine the properties of a unity-feedback system from the asymptotes of the open-loop frequency response. These charts utilize the following terminology (see Fig. 7-17):

$\left.\frac{C}{R}\right|_m$ — M_p, the maximum ratio of closed-loop frequency response

$\left.\frac{C}{R}\right|_p$ $\triangleq$ the peak value of the ratio of controlled variable (output) to reference variable (input) for a step input

$\frac{\omega_m}{\omega_c}$ $\triangleq$ the ratio of the frequency ω_m at which $\left.\frac{C}{R}\right|_m$ occurs to the frequency ω_c at which the straight-line approximation (asymptote) of the open-loop response is 0 decibels. (Note: 2 decilogs = 1 decibel.)

$\frac{\omega_t}{\omega_c}$ $\triangleq$ the ratio of ω_t, the lowest frequency of oscillation for a step input, to ω_c, the frequency at which the open-loop asymptote crosses 0 db (decibels).

$\omega_c t_p$ $\triangleq$ the asymptote crossover frequency ω_c times t_p, the response time from the start of the step function until $\left.\frac{C}{R}\right|_p$ occurs.

$\omega_c t_s$ $\triangleq$ the asymptote crossover frequency ω_c times t_s, the settling time from the start of the step function until the output continues to differ from the input by less than 5%.

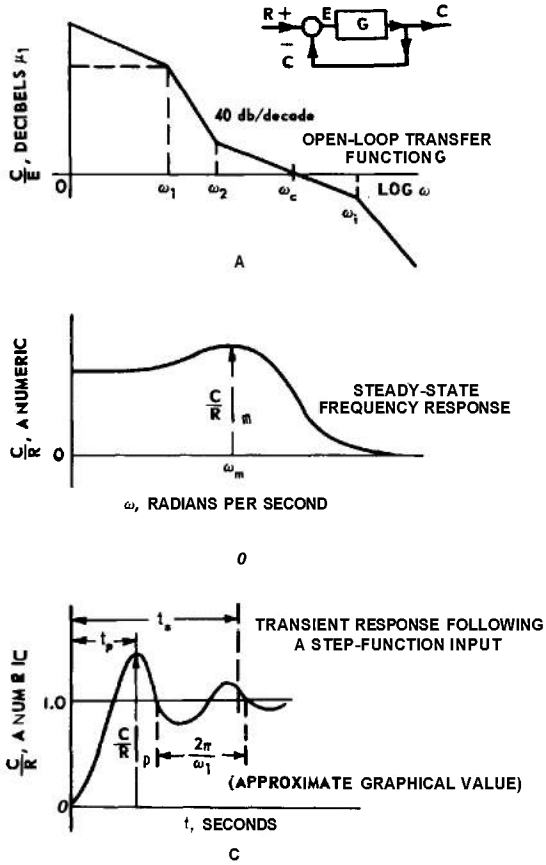


fig. 7-17 Sketches showing nomenclature used to describe various characteristics of servomechanism performance.

Reprinted with permission from *Servomechanisms and Regulating System Design*, Volume I, by H. Chestnut and R. W. Mayer, Copyright, 1961. John Wiley & Sons, Inc.

Additional definitions are given in Fig. 7-17. The charts correlating the quantities defined here are presented in Fig. 7-18 (sheets 1-18). It should be noted that these charts can be used either for analysis or for synthesis.

Biernson[11] presents an excellent method for determining the closed-loop poles of a system from the open-loop frequency response. If it is assumed that the asymptote crossover frequency ω_c of the open-loop frequency response occurs in (or near) a frequency region where the slope of the asymptote is -10dg/dec, then the following relations hold if $|G(j\omega)|$ is a monotonically decreasing function:

$$|G(j\omega)| \cong 1 \text{ for } \omega \cong \omega_c \qquad (7\text{-}43)$$

$$|G(j\omega)| >> 1 \text{ for } \omega << \omega_c \text{ (low-frequency range)} \qquad (7\text{-}44)$$

$$|G(j\omega)| << 1 \text{ for } \omega >> \omega_c \text{ (high-frequency range)} \qquad (7\text{-}45)$$

The *first* approximation to the location of the ***poles*** of the closed-loop transfer function $C(s)/R(s)$ is obtained from the following:

(a) The zeros of G(s) whose magnitudes are less than ω_c (low-frequency zeros)

(b) The poles of G(s) whose magnitudes are greater than ω_c (high-frequency poles)

(c) A pole at $s = -\omega_c$

For real or complex closed-loop poles which are far from ω_c in magnitude, the shift from the first approximation of these poles to their actual location can be calculated by a successive-approximation method which converges more rapidly the further the poles are from ω_c. If a closed-loop pole is approximated by a ***low-frequency zero*** of $G(s)$, then the true location of the closed-loop pole s_1 can be determined by successively evaluating

$$(\delta_a)^n \cong \left[\frac{(s - s_a)^n}{G(s)}\right]_{s=s_1} \quad \text{for } |s_a| < \omega_c, \text{ and } s_1 \cong s_a \qquad (7\text{-}46)$$

where n is the order of the open-loop zero, s_a is the location of the open-loop zero, and δ_a is the shift from the open-loop zero to the closed-loop pole, i.e., $\delta_a = s_1 - s_a$.

If a closed-loop pole is approximated by a ***high-frequency*** pole of $G(s)$, then the true location of the closed-loop pole s_2 can be determined by evaluating

$$(\delta_b)^n \cong -[(s - s_b)^n G(s)]_{s=s_2} \quad \text{for } |s_b| > \omega_c, \text{ and } s_2 \cong s_b \qquad (7\text{-}47)$$

where n is the order of the open-loop pole, s_a is the location of the open-loop pole, and δ_b is the shift from the open-loop pole to the closed-loop pole, i.e., $\delta_b = s_2 - s_b$.

For closed-loop poles near ω_c, a graphical **procedure**[11] is recommended since the convergence of the numerical method employing Eqs. **(7-46)** and **(7-47)** is either slow or nonexistent. The graphical procedure involves plots of G(s) for $s = -\sigma + j\omega$ (along axes other than the imaginary axis). Because the graphical procedure tends to be somewhat lengthy, it will not be given here.

Example. The open-loop transfer function of a unity-feedback system is given by

$$G(s) = \frac{K_o(s + \omega_2)(s + \omega_3)}{s(s + \omega_1)^2(s + \omega_4)^2} \qquad (7\text{-}48)$$

where

$\omega_1 = 0.04$

$\omega_2 = 0.2$

$\omega_3 = 1$

$\omega_4 = 16$

K_o is a proportionality constant whose value is to be determined. The asymptotes of this function are sketched in Fig. 7-19. The crossover frequency ω_c is chosen as the geometric mean of ω_3 and ω_4 since this particular choice tends to produce the lowest closed-loop M_p. Near ω_c, the asymptote is given by

$$\text{Asymp } |G| = \left|\frac{K_o}{s\omega_4^2}\right|$$

At the crossover frequency, ω_c, therefore,

$$\frac{K_o}{\omega_c\omega_4^2} = 1$$

or

$$K_o = \omega_c\omega_4^2 = \sqrt{\omega_3\omega_4}\,\omega_4^2 = 4\omega_4^2$$

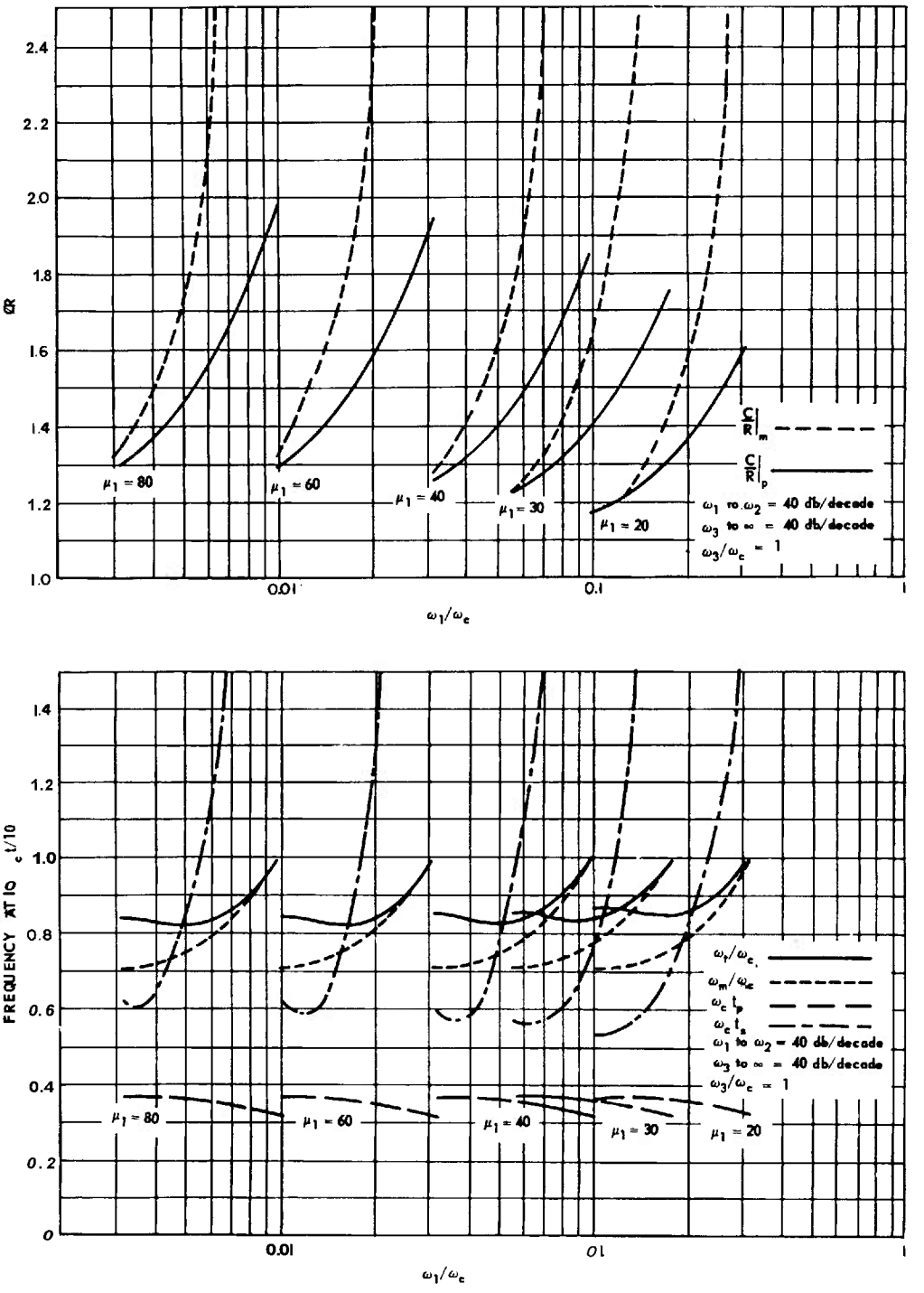


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 1 of 78)

Reprinted with permission from *Servomechanisms and Regulating System Design.* Volume I. by H. Chestnut and R. W. Mayer, Copyright, 1961, John Wiley & Sons, Inc.

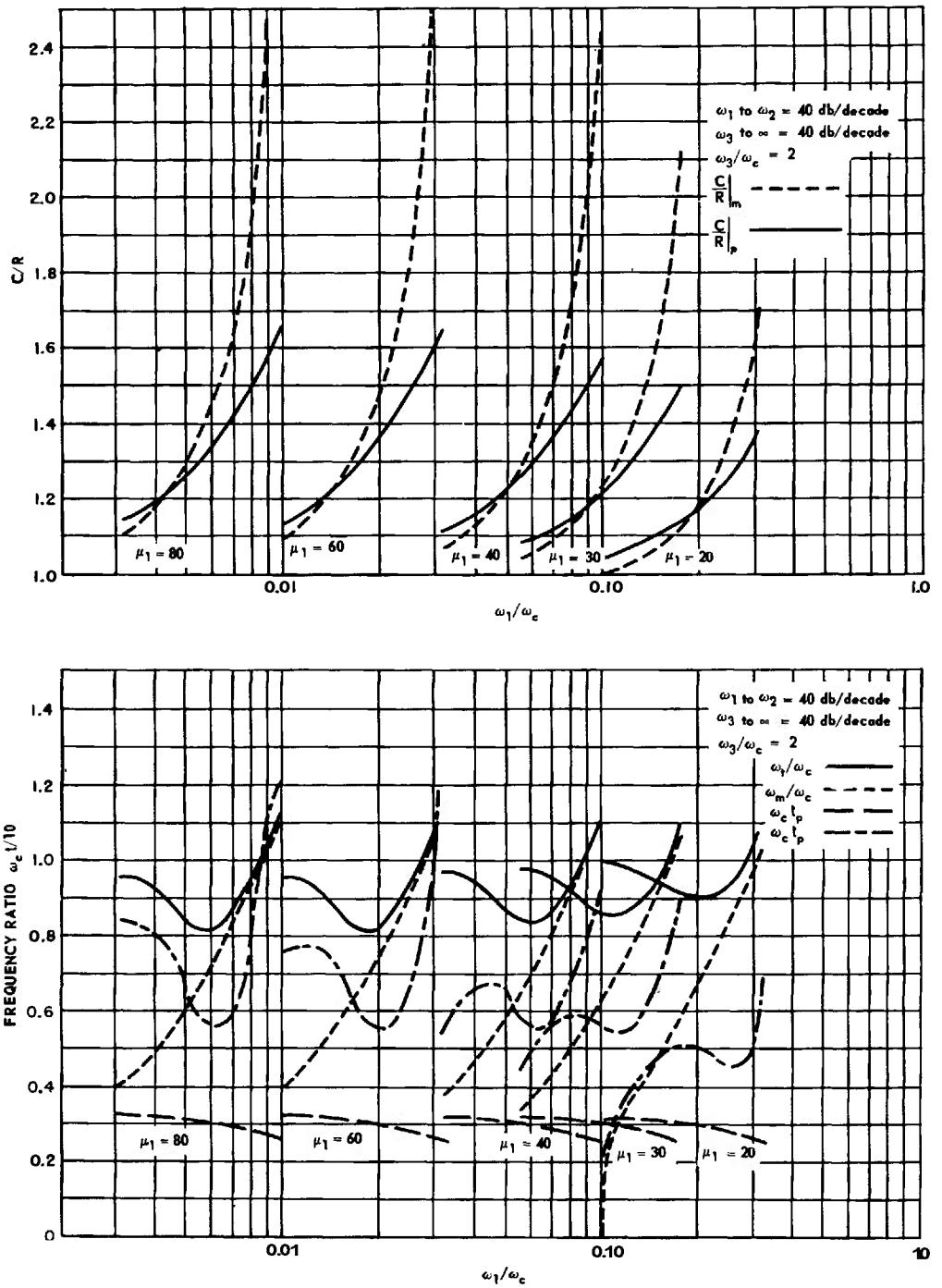


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 2 of 18)

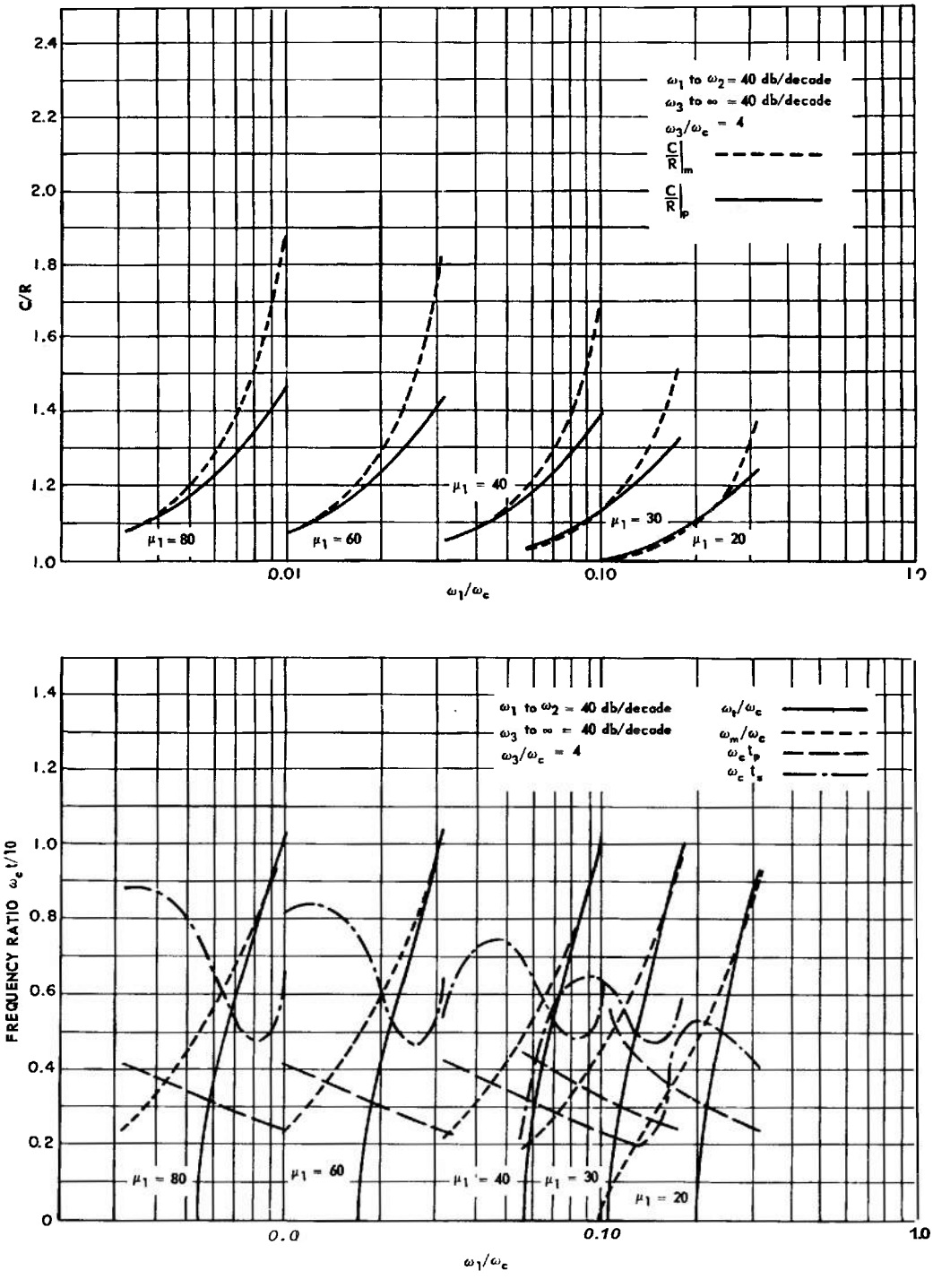


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 3 of 78)

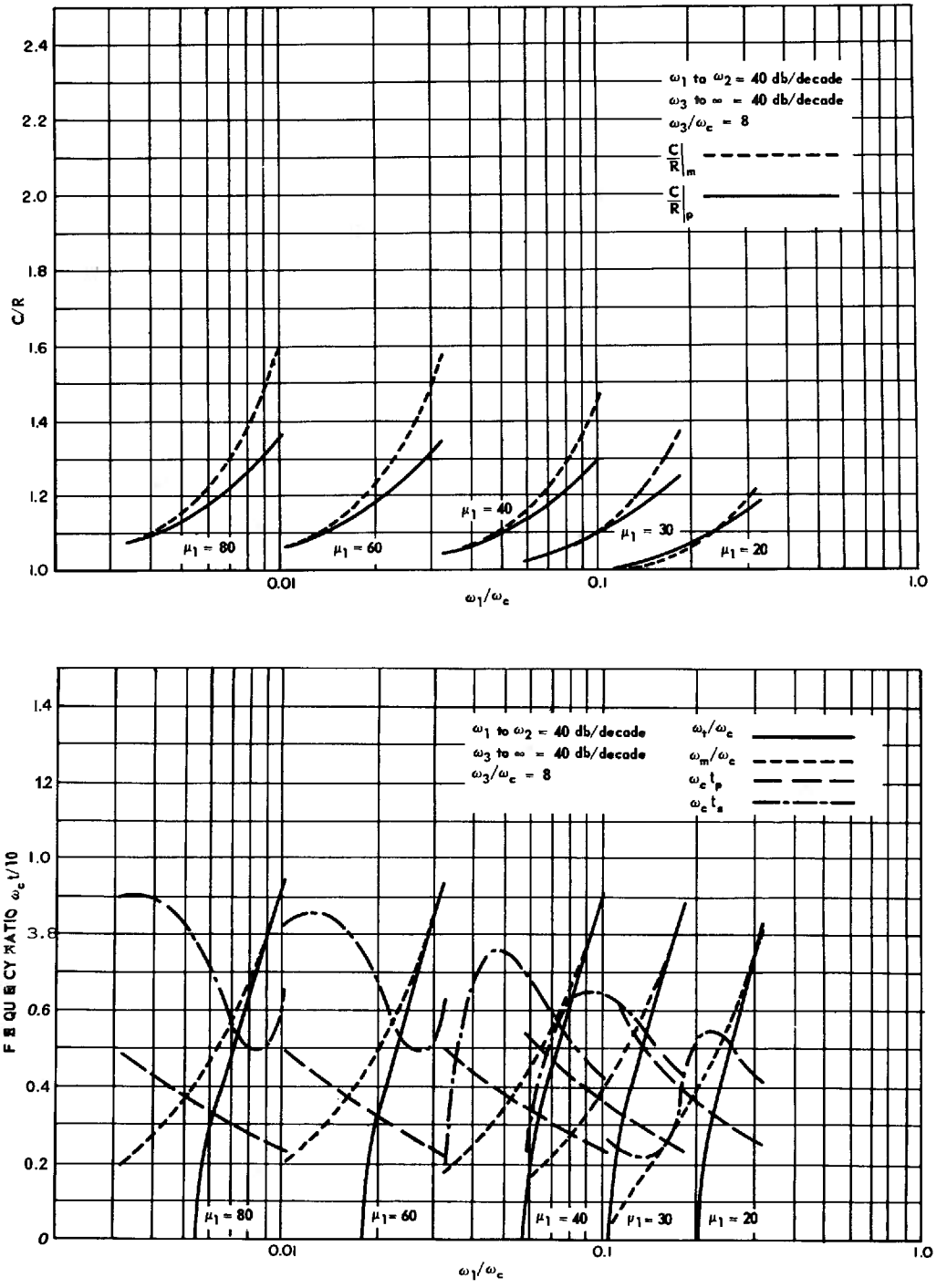
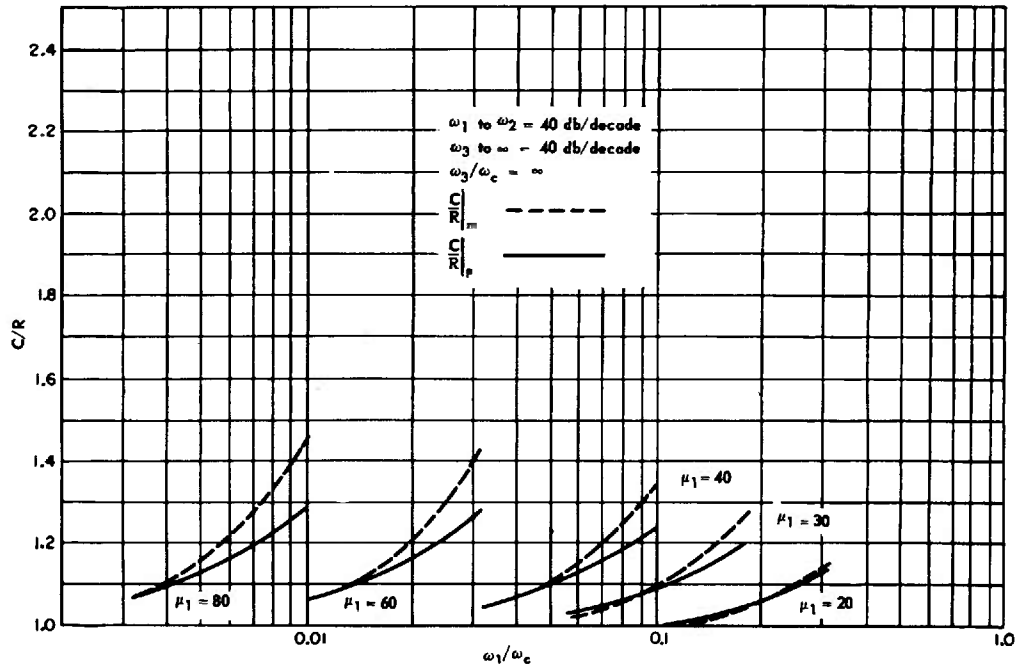


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 4 of 18)



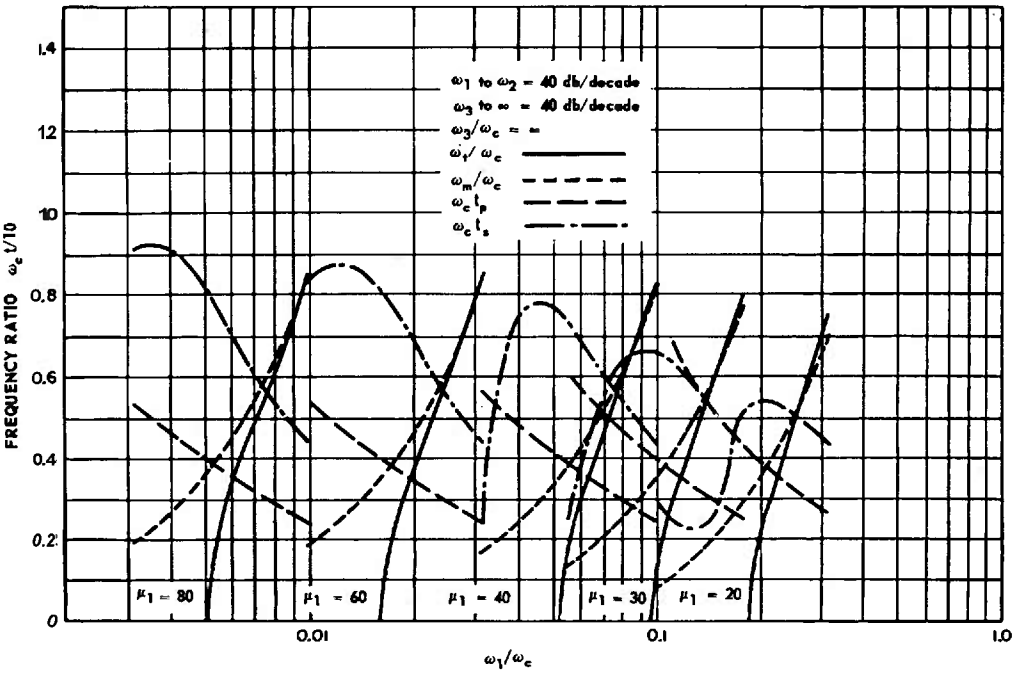


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 5 of 78)

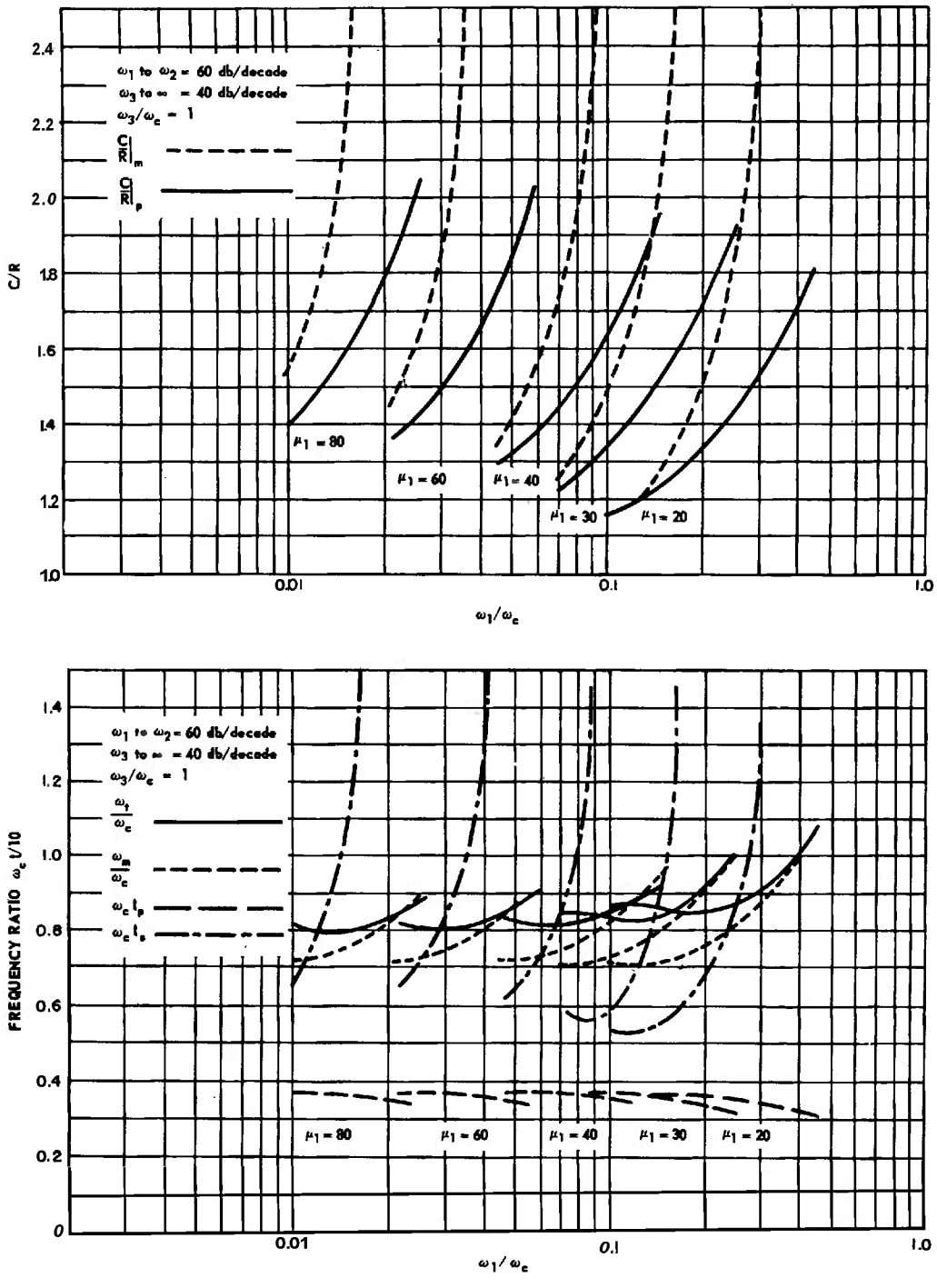


Fig. 7-78 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 6 of 78)

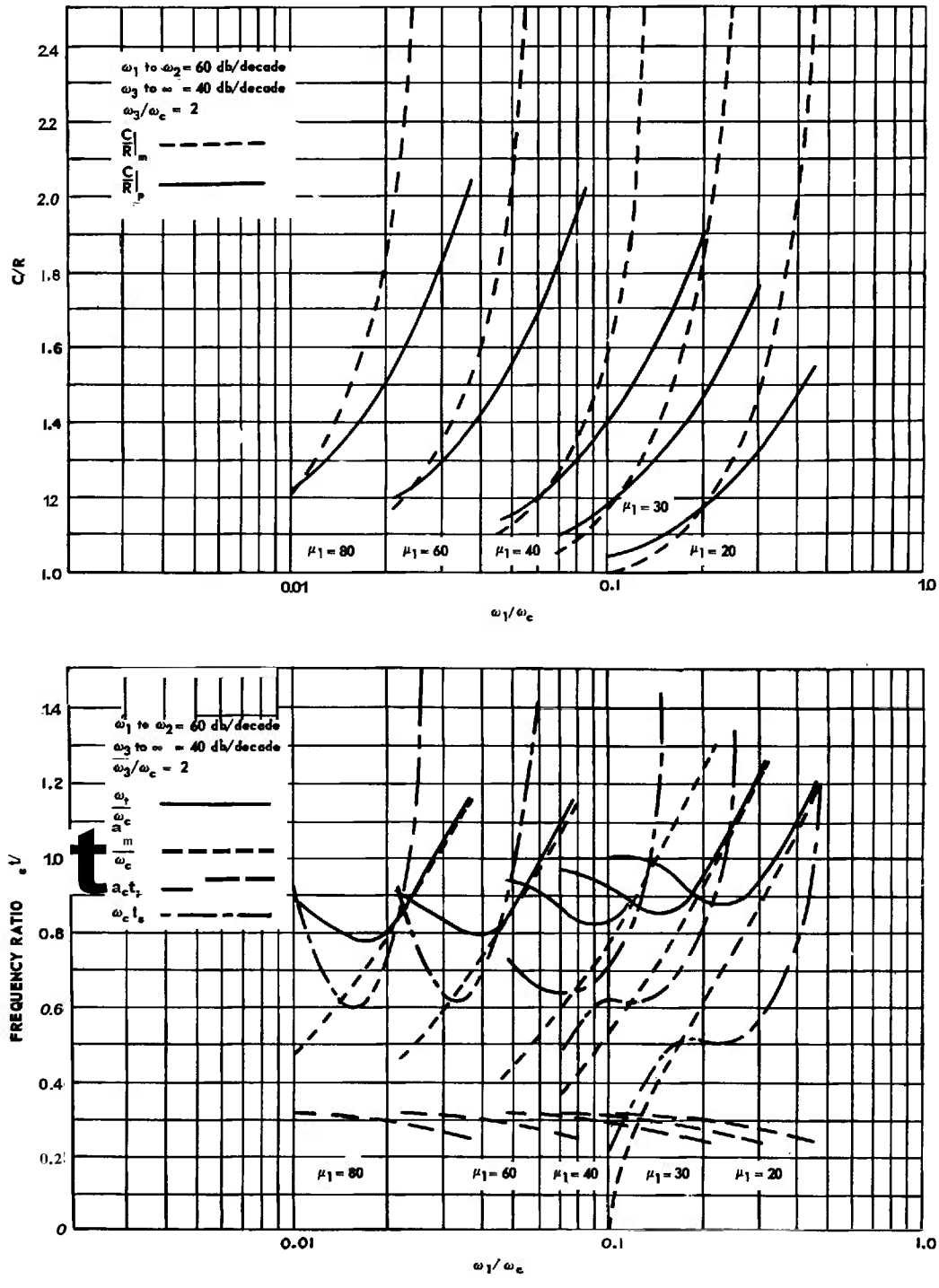


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a fundion of ω_1/ω_c. (Sheet 7 of 78)

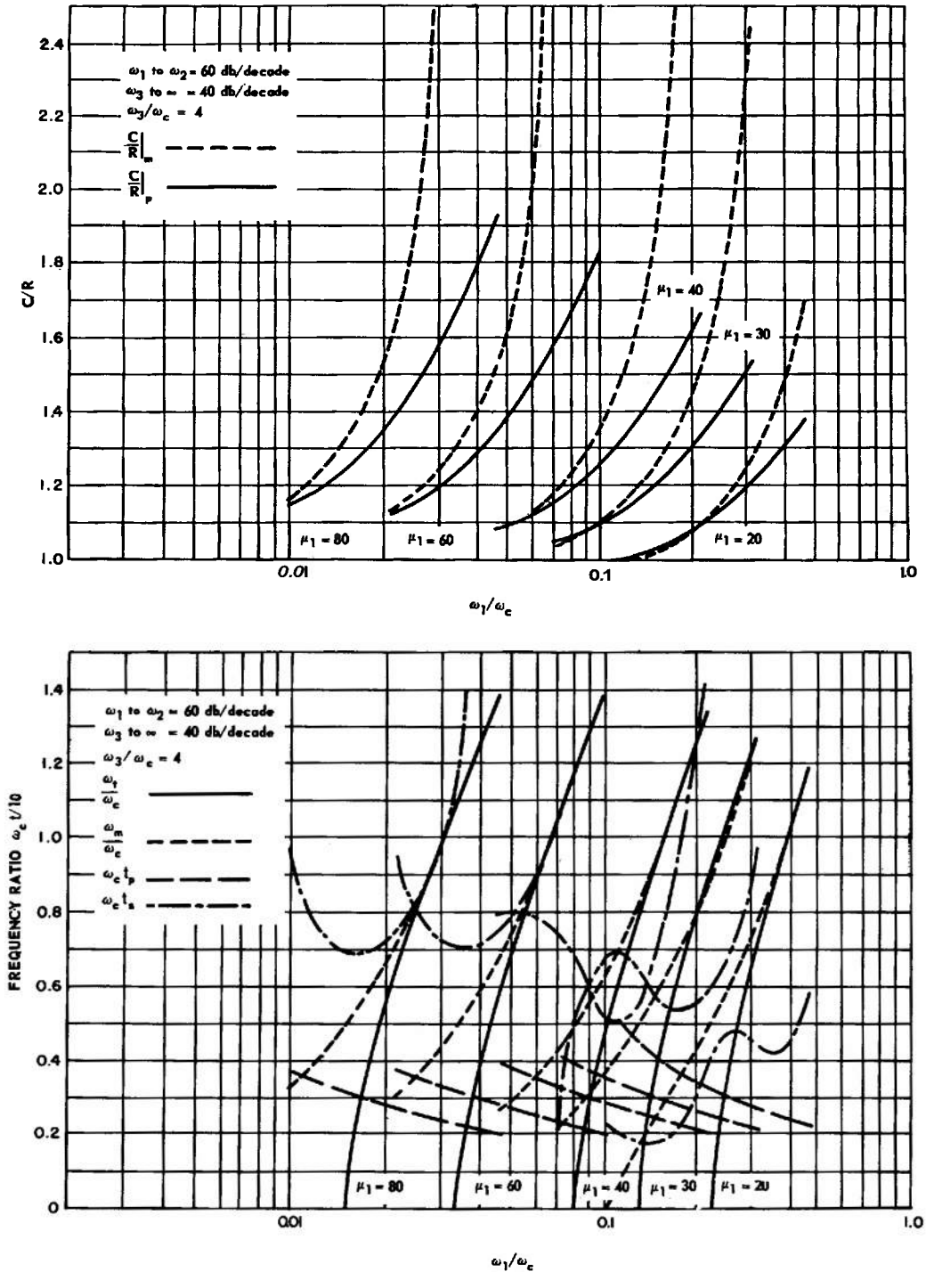


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 8 of 78)

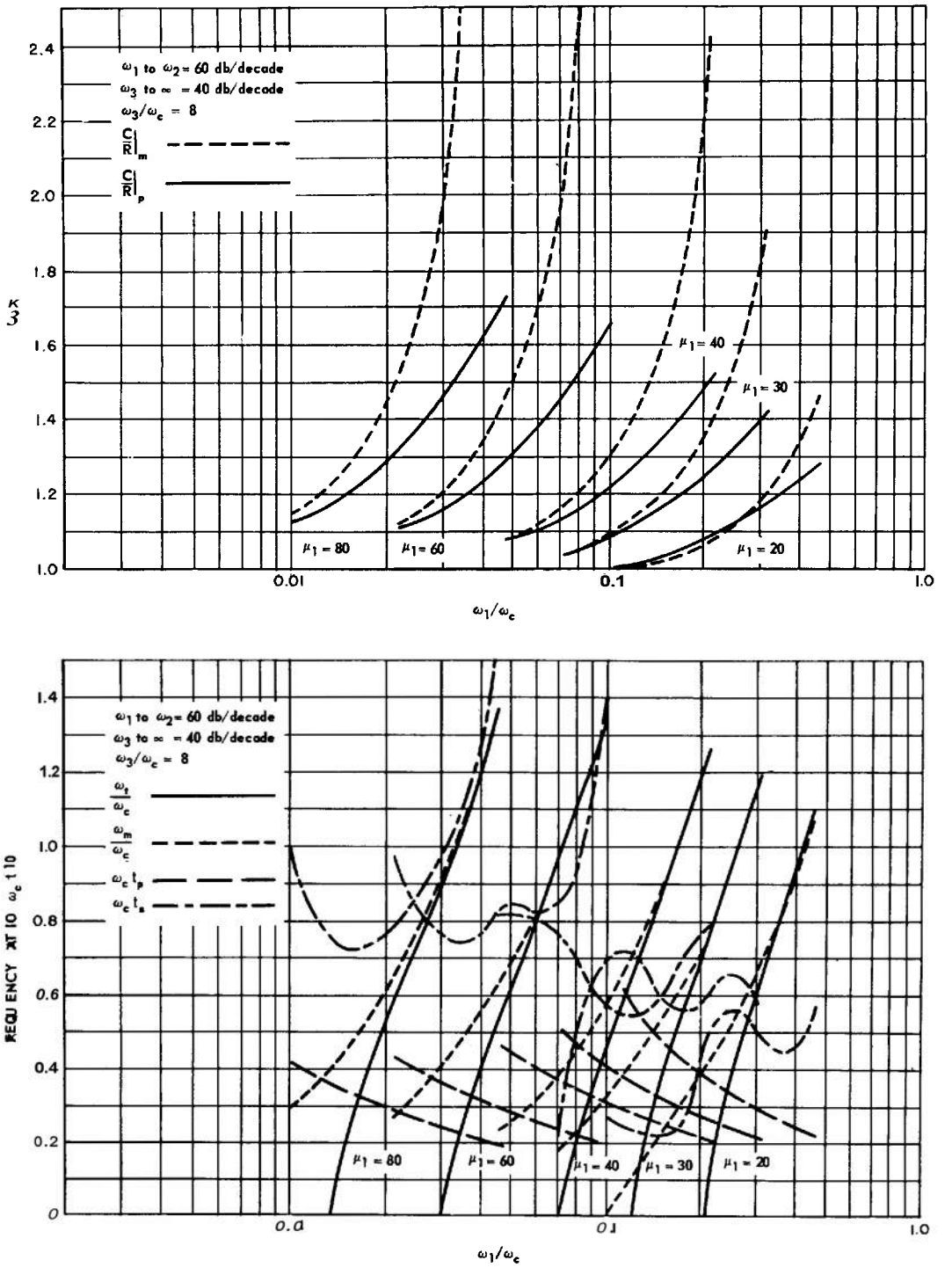


fig. 7-*18* Comparison of steady-state frequency response characteristics *and* transient response following a step function of input *as* a function of ω_1/ω_c. (Sheet *9* of *18)*

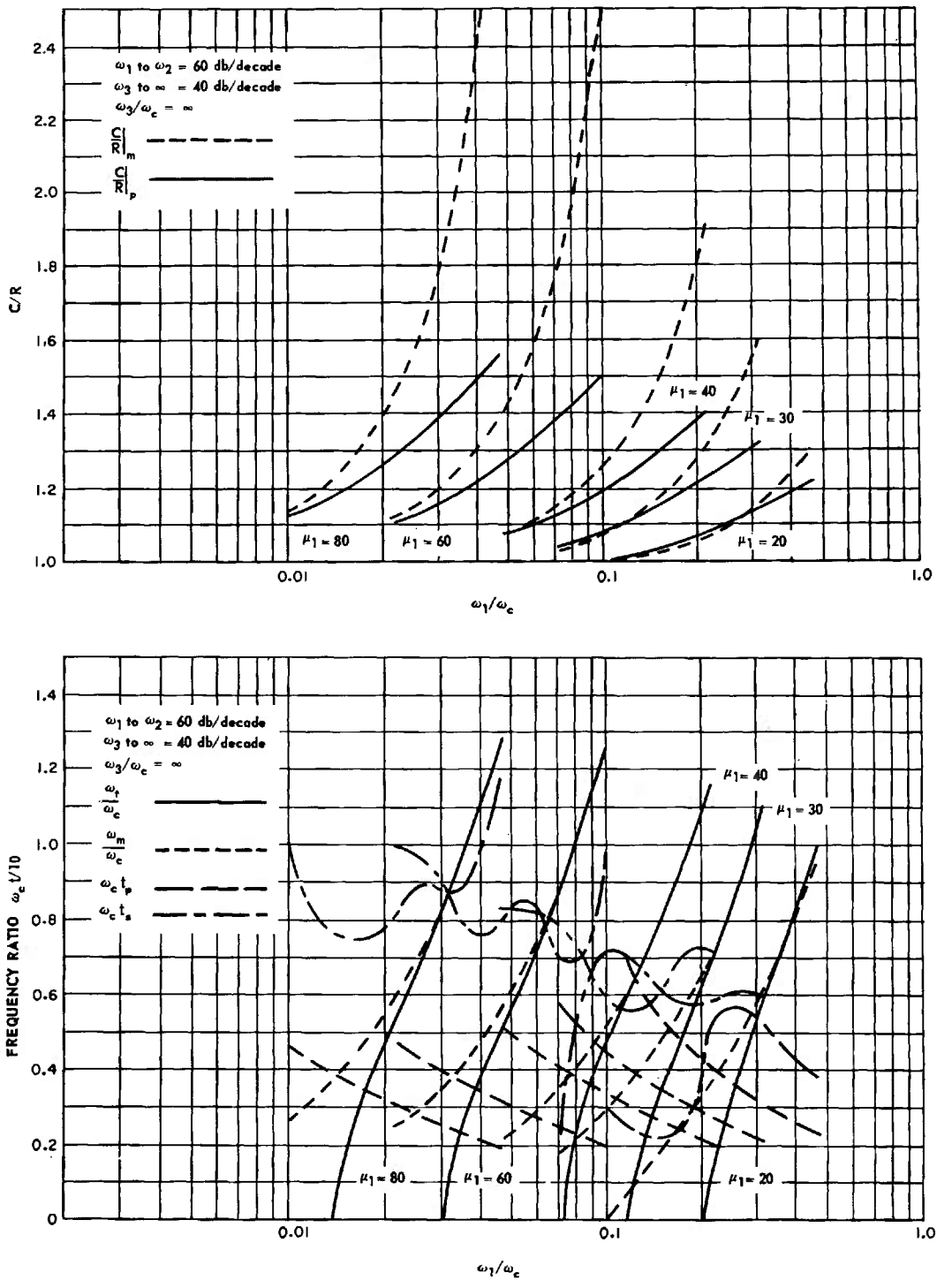


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 10 of 18)

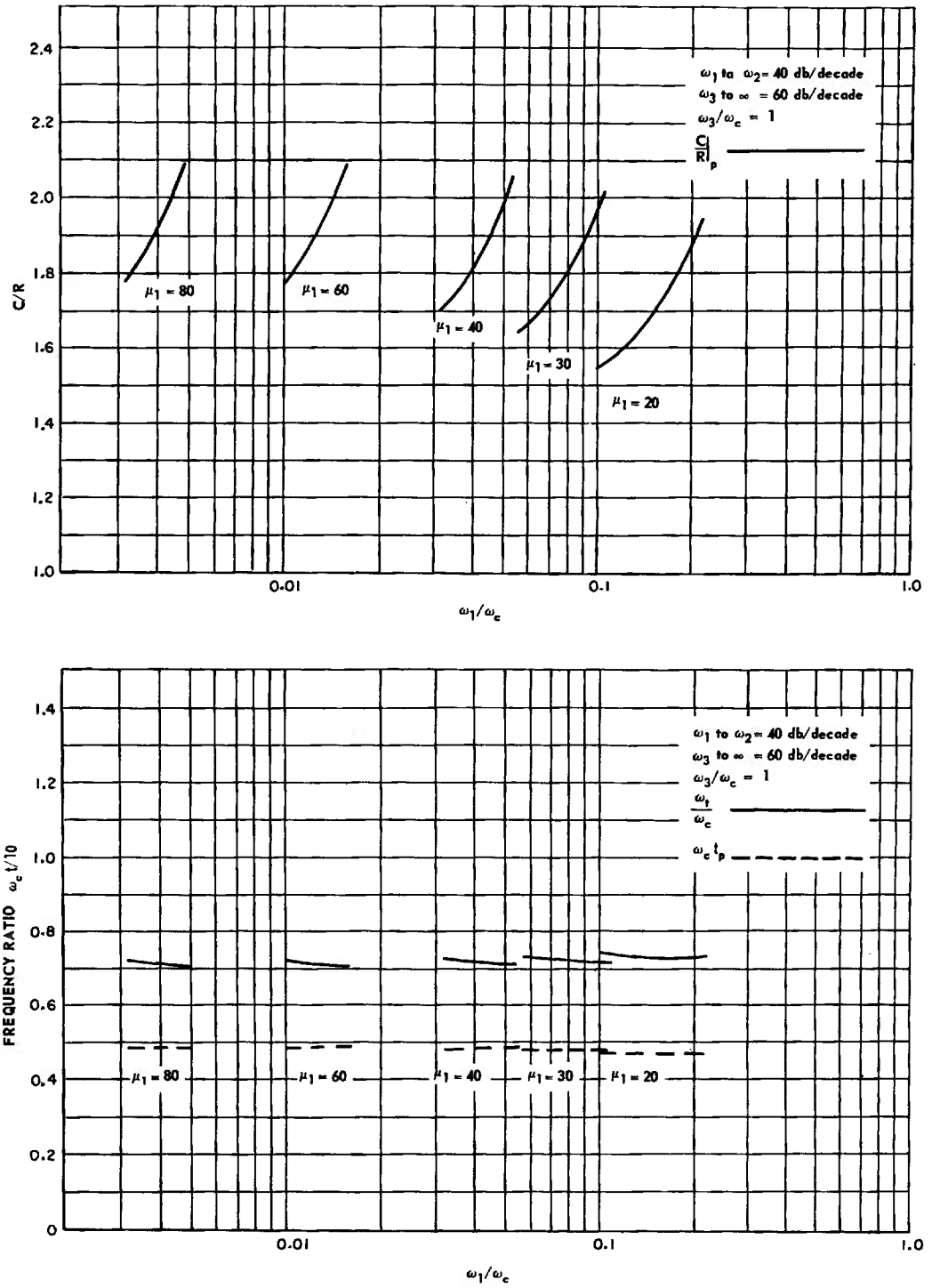


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step fundion of input as a function of ω_1/ω_c. (Sheet 11 of 78)

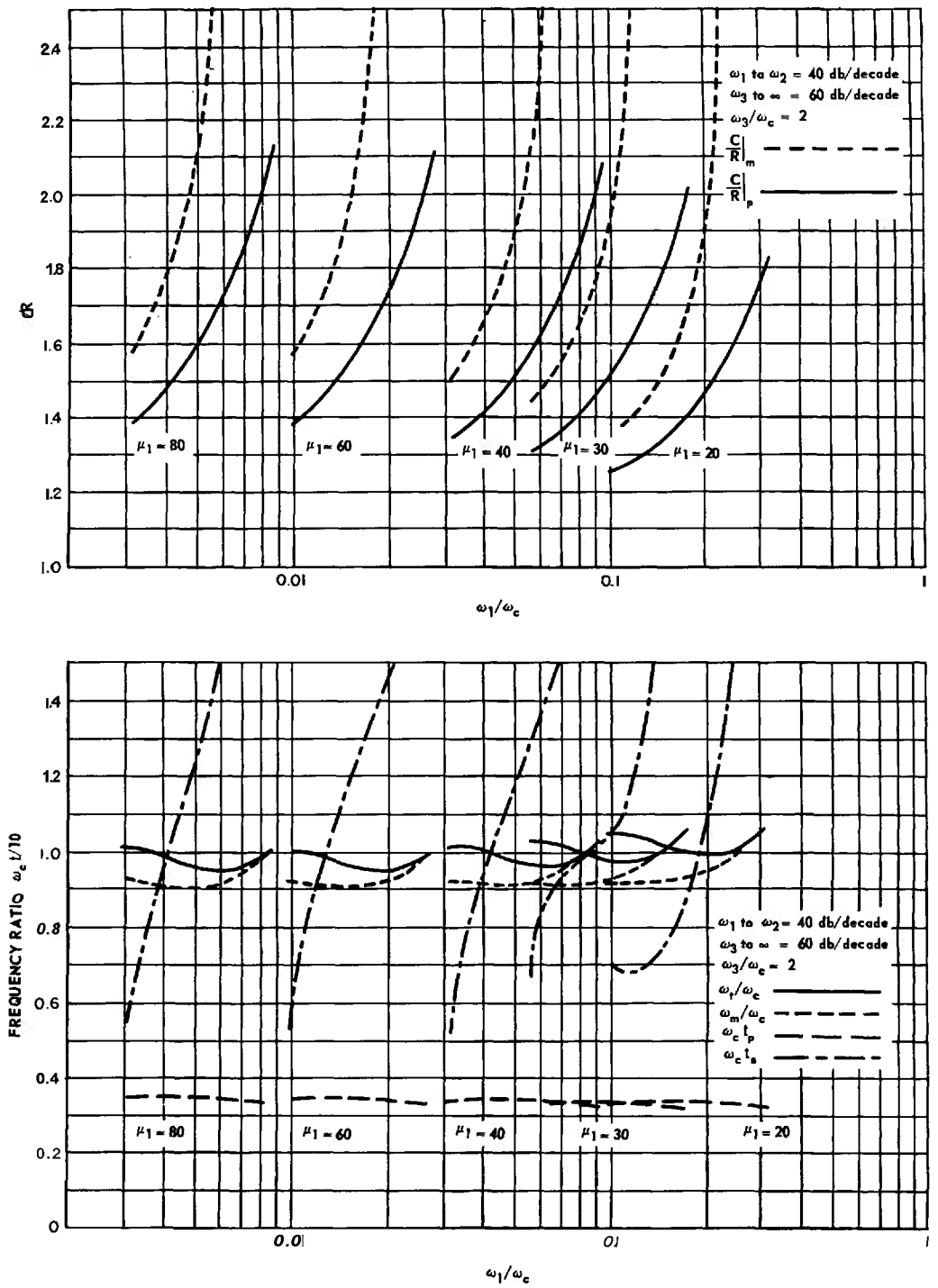


Fig. 7-18 Comparison of steady-state frequency response characteristics and fransient response following a step function of input as a function of ω_1/ω_c. (Sheet 12 of 18)

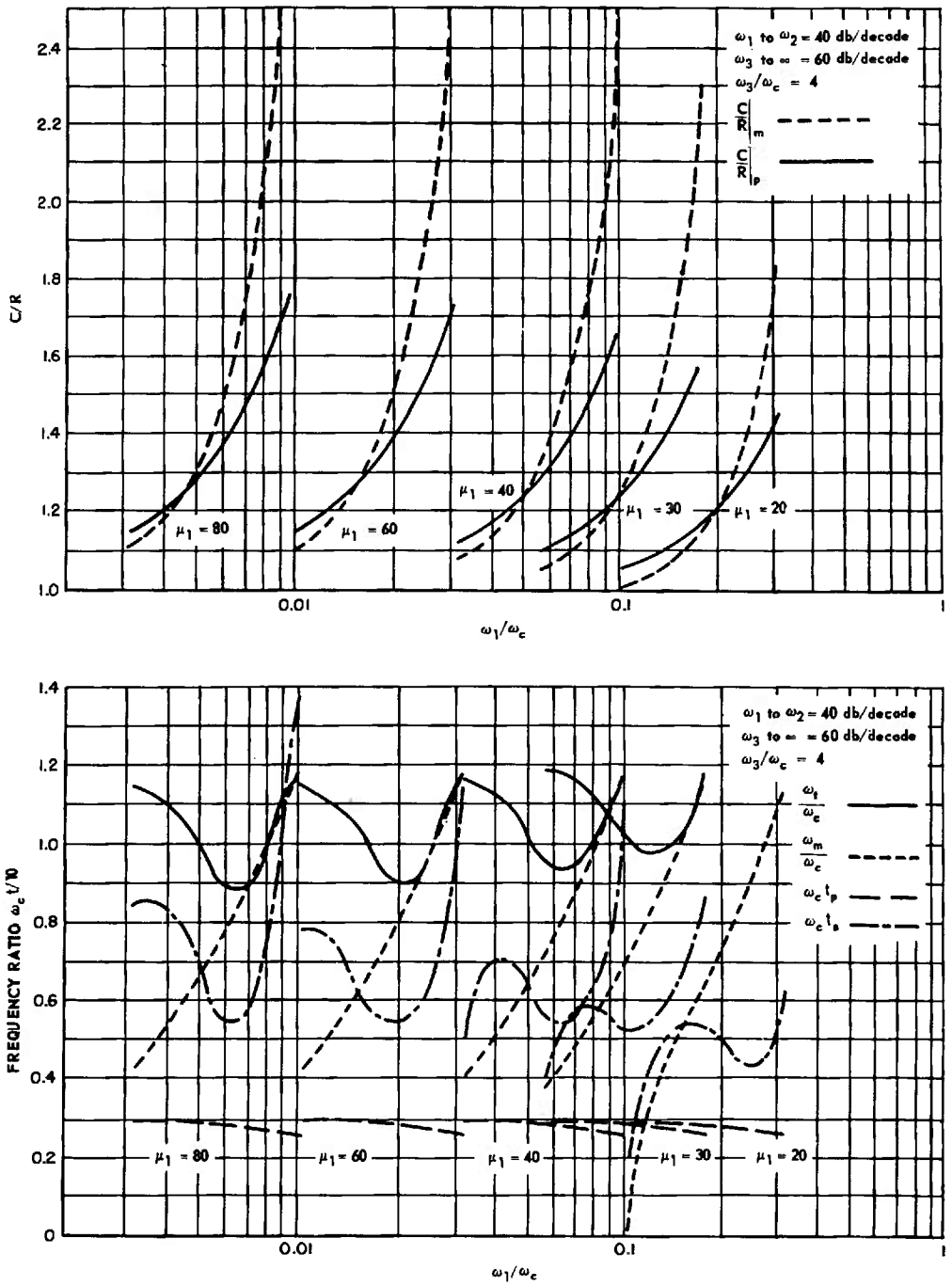


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 13 of 18)

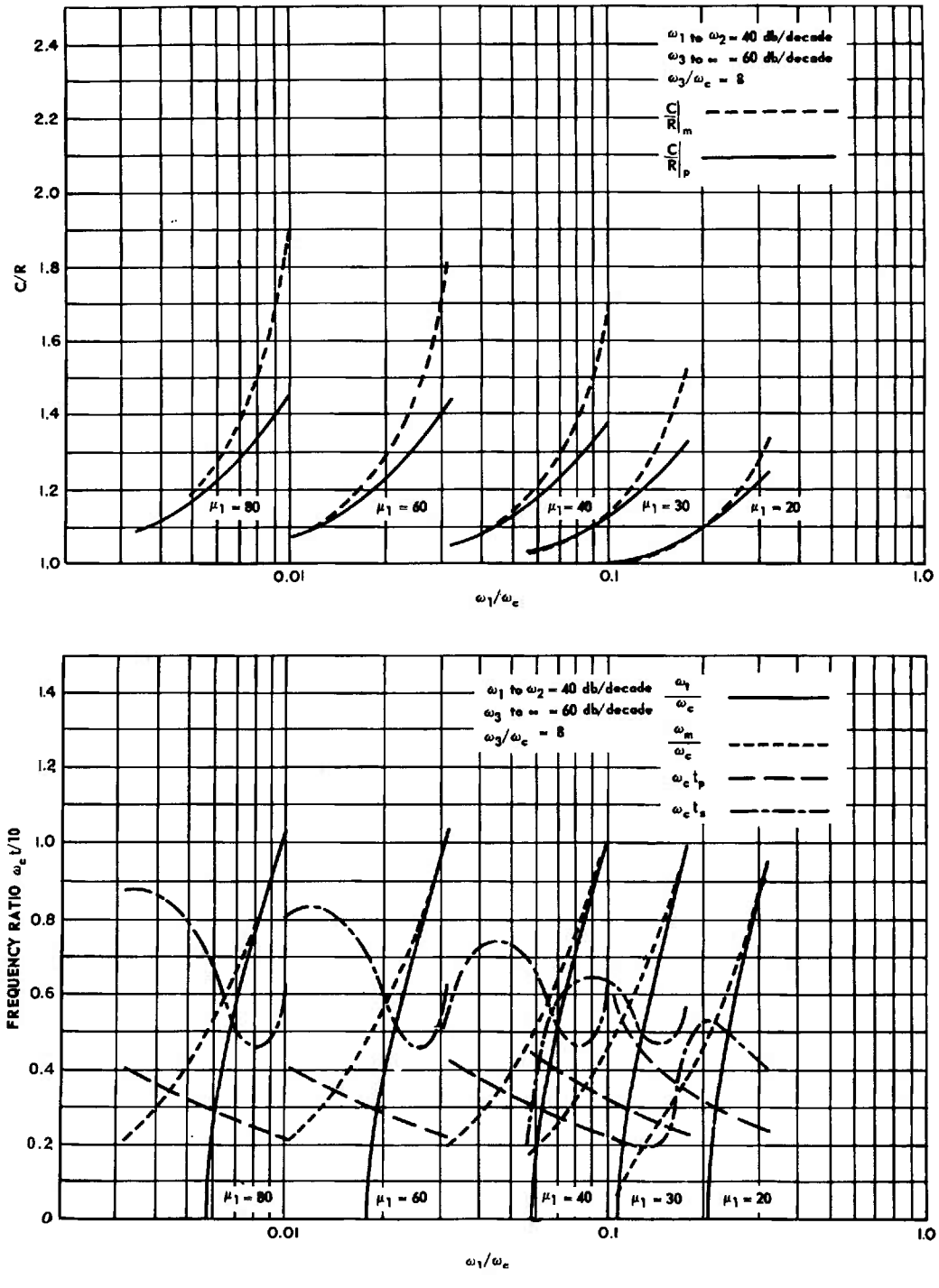


Fig. 7-18 Comparison of *steady-state* frequency response characteristics *and* transient response following a step function of input as a function of ω_1/ω_c. (Sheet *14* of *18)*

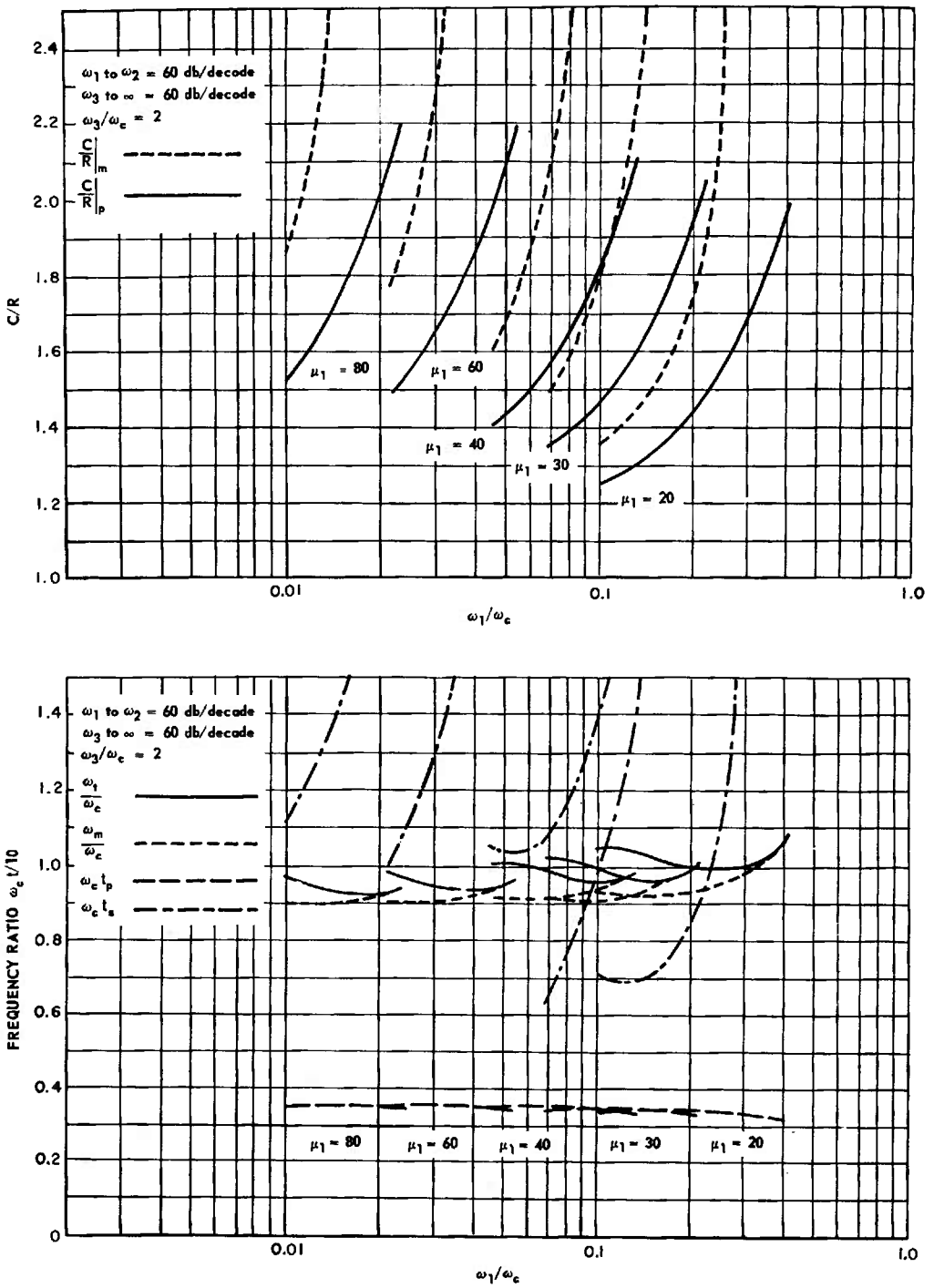


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 15 of 18)

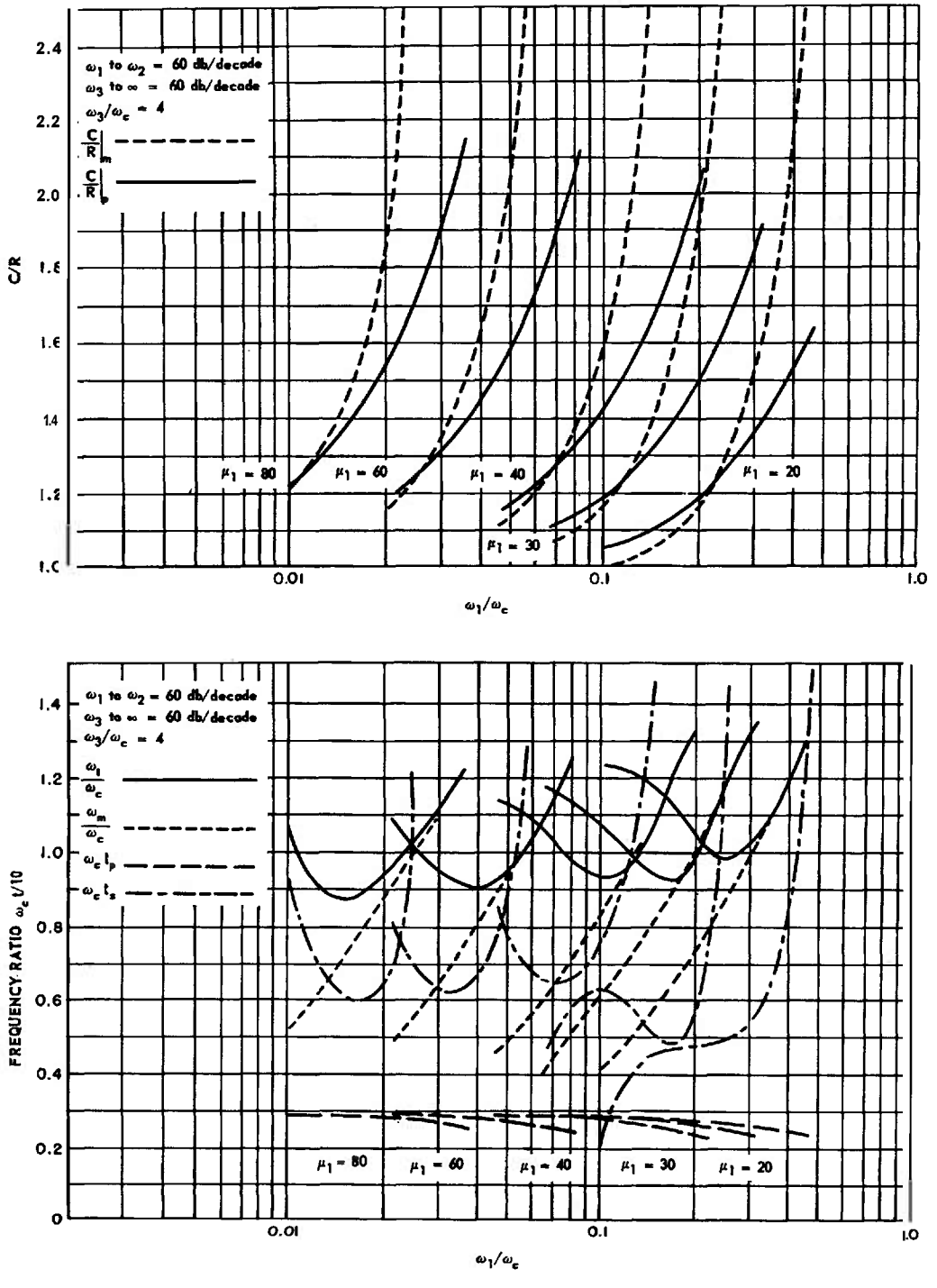


fig. 7-18 Comparison of steady-state frequency response characteristics and fransient response following a step function of input as a function of ω_1/ω_c. (Sheet 76 of 78)

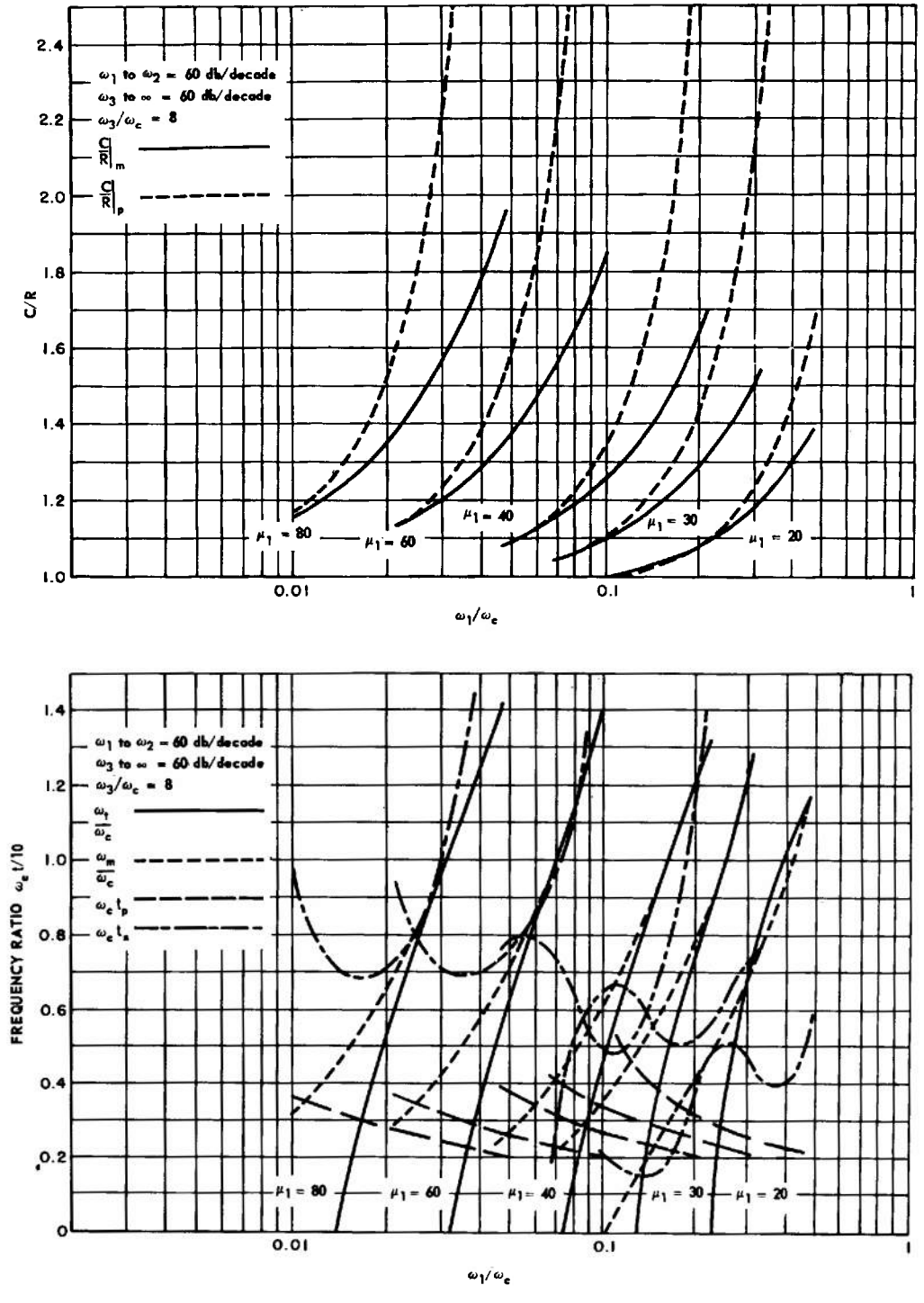


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c. (Sheet 77 of 78)

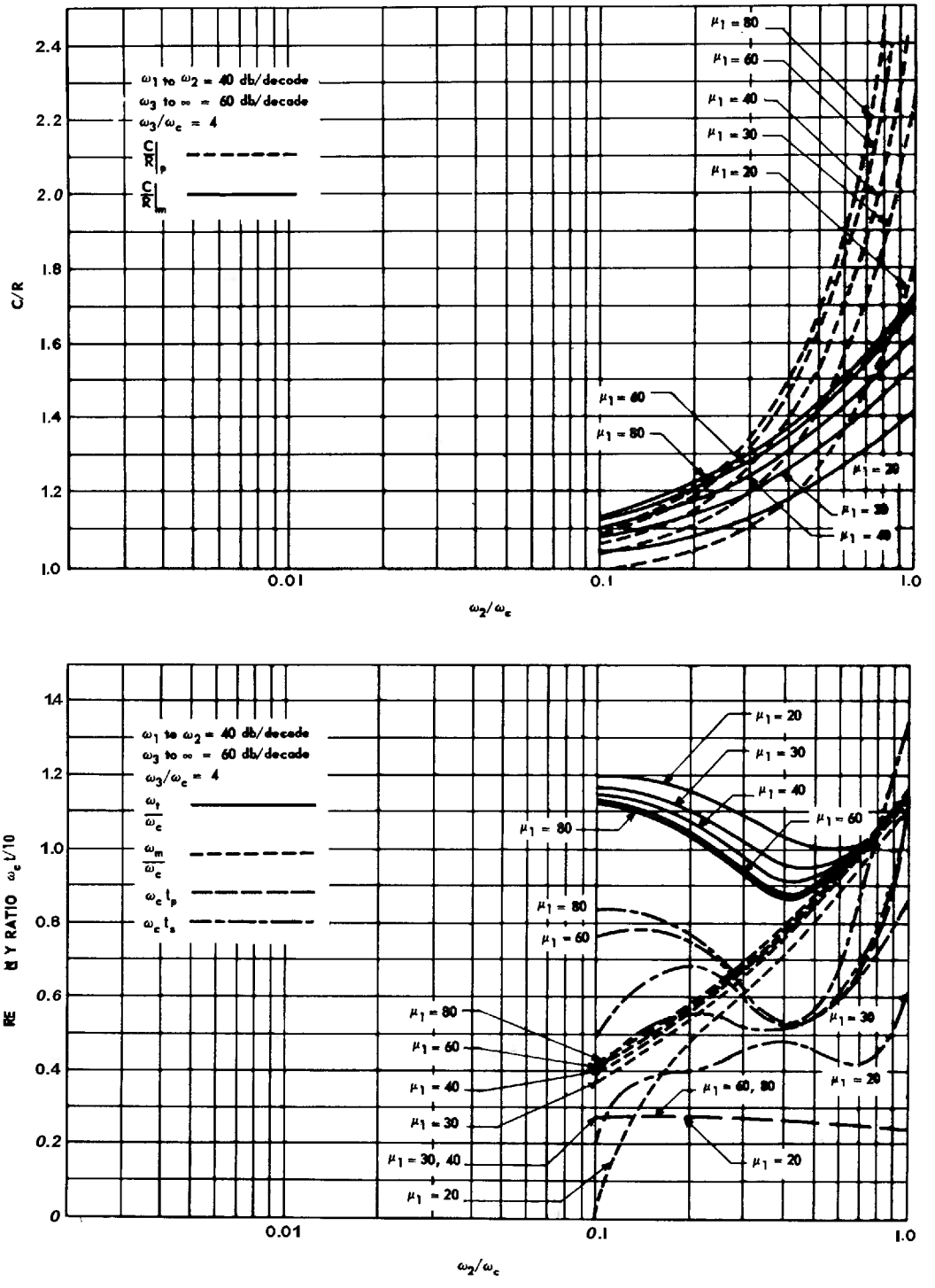


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step fundion of input as a function of ω_r/ω_c. (Sheet 78 of 78)

Inspection of Fig. 7-19 and Eq. (7-48) shows that the low-frequency zeros of G(s) occur at $-\omega_2$ and $-\omega_3$ and that the high-frequency poles of $G(s)$ consist of a double pole at $-\omega_4$. Therefore, the poles of the first approximation to $C(s)/R(s)$ occur at $-\omega_2$ (single-order), $-\omega_3$ (single-order), and $-\omega_4$ (double-order), plus one pole at $-\omega_c$. The zeros of $C(s)/R(s)$ are the zeros of $G(s)$, $-\omega_2$ and $-\omega_3$. As a result, the first approximation to $C(s)/R(s)$ is

$$\frac{C(s)}{R(s)} \cong \frac{\omega_c \omega_4^2}{(s+\omega_c)(s+\omega_4)^2} \cong \frac{K_o \quad \text{(zeros of G(s) above } \omega_c)}{(s+\omega_c) \quad \text{(poles of G(s) above } \omega_c)} \tag{7-49}$$

The first approximation to $E(s)$ is found to be

$$\frac{E(s)}{R(s)} \cong \frac{s(s+\omega_1)^2}{(s+\omega_c)(s+\omega_2)(s+\omega_2)} \cong \frac{1 \quad \text{(poles of G(s) below } \omega_c)}{(s+\omega_c) \quad \text{(zeros of G(s) below } \omega_c)} \tag{7-50}$$

The approximate factors of $1+G(s)$ are given by

$$1+G(s) \cong (s+0.2)(s+1)(s+4)(s+16)^2 \tag{7-51}$$

To evaluate the shifts from the approximate factors given in Eq. (7-51), Eqs. (7-46) and (7-47) are used. The numerical form of $G(s)$ is

$$G(s) = \frac{4 \times 16^2 (s+0.2)(s+1)}{s(s+0.04)^2(s+16)^2} \tag{7-52}$$

Since $\omega_c = 4$, the numerical method should be tried for the factors $(s+0.2)$, $(s+1)$, and $(s+16)$. The shift from the approximate pole at $s = -0.2$ to the true pole is given by

$$\delta_2 \cong -\left.\frac{(s+0.2)}{G(s)}\right]_{s=-0.2} = 0.0016 \tag{7-53}$$

Since this quantity is small, the true pole lies at

$$s = -0.2 + 0.0016 = -0.1984$$

For the approximate pole at $s = -\omega_3 = -1$, the shift to the true pole is given by

$$\delta_3 \cong -\left.\frac{(s+1)}{G(s)}\right]_{s=-1} = -0.253 \tag{7-54,}$$

This quantity is reasonably large so that a second approximation to δ_3 is made by evaluating the right side of Eq. (7-54) at $s = (-1) + (-0.253)$, instead of at $s = -1$, yielding

$$\delta_3 \cong \left.\frac{-(s+1)}{G(s)}\right]_{s=-1.253} = -0.373 \tag{7-55}$$

The shift is still not too well approximated since the change from Eq. (7-54) to (7-55) is significant. The third approximation to δ_3 is obtained by evaluating Eq. (7-55) at $s = (-1) + (-0.373)$, instead of at $s = -1.253$, yielding

$$\delta_3 \cong \left.\frac{-(s+1)}{G(s)}\right]_{s=-1.373} = -0.435 \tag{7-56}$$

Since the original approximate closed-loop pole at $s = -1$ was fairly close to the crossover frequency $\omega_c = 4$, it is to be expected that the process of determining the shift δ_3 will converge fairly slowly. The succeeding approximations to the value of the pole are as follows:

-1.468, -1.485, -1.496, -1.503, -1.507, -1.509, and finally -1.510.

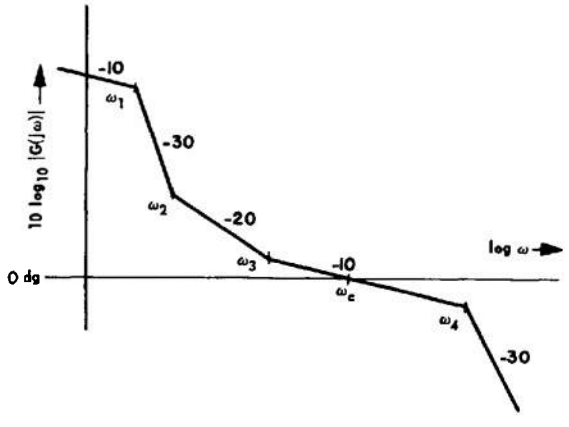


Fig. 7-19 **Sketch of open-loop asymptote function.**

For the approximate double-order pole at $s = -16$, the shift is given by

$$(\delta_4)^2 \cong -(s+16)^2 G(s)\Big]_{s=-16} = 59.6 \tag{7-57}$$

or

$$\delta_4 = \pm 7.7$$

The original double-order pole splits in two, one pole moving toward ω_c and the other pole moving away from ω_c. Only the pole moving away from ω_c can be determined by the numerical method. The second approximation to the shift of this pole is $s = (-16) + (-7.7) = -23.7$. The third approximation to the shift of this pole is obtained by evaluating Eq. (7-57) at s = −23.7, instead of at s = −16, yielding

$$(\delta_4)^2 \cong -(s+16)^2 G(s)\Big]_{s=-23.7} = \pm 6.4 \tag{7-58}$$

or $s = -22.4$ is the third approximation to the pole. Keeping track of the negative shift in this case, the succeeding values of the pole are found to be −22.6 and finally −22.56.

Thus, three of the five "exact" factors of $1 + G(s)$ are $(s + 0.1984)$, $(s + 1.51)$, and $(s + 22.56)$. Dividing these factors out yields the remaining complex poles at $s = -4.35 \pm j3.3$. The "exact" close-loop response C/R is therefore

$$\frac{C(s)}{R(s)} = \frac{1014\,(s+0.2)\,(s+1)}{(s+0.1984)\,(s+1.51)\,(s+22.56)\,(s^2+8.69s+30)} \tag{7-59}$$

Although the detail with which this example has been presented may make the procedure seem laborious, actually it is extremely rapid even when the rate of convergence of the successive approximations is relatively slow. Note, also, that any desirable degree of accuracy can be maintained.

In other papers [23,25], Biernson gives a very good summary of approximate relations between the open-loop frequency response and the closed-loop transient response. If ω_{cm} is defined as the frequency at which the magnitude of $G(j\omega)$ crosses the 0-dg line (magnitude crossover frequency), then

(a) The ***rise time*** is approximately $1/\omega_{cm}$, where the rise time is defined as the time for the output response to a unit-step input to reach 0.63.

(b) The ***peak error for a unit-ramp*** input is approximately $1/\omega_{cm}$.

(c) The ***peak output response for a unit-impulse*** input is approximately ω_{cm}.

(d) The ***peak overshoot in the output response*** to a unit-step input is best determined from M_p by means of Figs. 7-14 and 7-15.

(e) The ***settling time*** t_s is approximated by the settling time of the equivalent second-order system unless $G(s)$ has low-frequency zeros produced by integral networks.

(f) If a first-order lag network (integral network) has been used to compensate the system, the peak overshoot of the ***output*** response to a unit-step input will be increased. If T_c is the time constant of the lag network, an additional transient term Ae^{-t/T_c} is added to the step response, where $A \cong 1/T_c\omega_c$ and ω_c is the ***asymptote*** crossover frequency.

(g) If an integral network is added to a system, the rate of decay of the ***error*** response to a unit-ramp input response is determined by the time constant T_c of the integral network. This response can be sketched from the following considerations:

(1) The response initially rises at the same rate as the input.

(2) The peak of the response is $1/\omega_{cm}$.

(3) If ω_c is the asymptote crossover frequency, the tail of the response is approximately the tail of an exponential with time constant T_c starting from $1/\omega_c$ at $t = 0$ and falling to $1/K_v$ at $t = \infty$, where K_v is the velocity constant.

(h) The ***maximum*** time delay by which the output response to a unit-ramp input lags the input ramp is approximately equal to the rise time (0.63-value) of the output response to a unit-step input.

7-2 ERROR COEFFICIENTS[17,23,27,31,32]

Paragraph 7-1 discusses the performance of feedback-control systems in terms of transient response and frequency response. Both of these views are intimately connected and stem from the impulse-response and convolution-integral description which forms the basis for all performance-evaluation methods. Unfortunately, it is generally true that the evaluation of performance by any **of** the foregoing methods is very laborious when the input is an arbitrary (but definable) time function. The error-coefficient method is a technique which aids the designer in such a case. Paragraph **3-6** shows that the error response of a system to a specified input can be expressed in terms of the input, its derivatives, and a set of ***error Coefficients*** derivable from the transfer function of the system. The expression for the error response is

$$e(t) = e_0 r(t) + e_1 r'(t) + e_2 r''(t) + \dots \quad \textbf{(7-60)}$$

where the error coefficients e_0, e_1, e_2, . . . are the coefficients of the Maclaurin series expansion of the error-to-input transfer function $E(s)/R(s)$, i.e.,

$$\frac{E(s)}{R(s)} = e_0 + e_1 s + e_2 s^2 + \dots \quad \textbf{(7-61)}$$

This expansion is valid everywhere except where the input or any of its derivatives are discontinuous. For practical purposes, only a few terms of the expansion are used to evaluate the error response. However, the expansion cannot be used near points of discontinuity of r, r', r'', ... if accurate results are sought. Thus, for example, if a step discontinuity occurs in the input $r(t)$, the expansion is invalid for a time interval extending from the instant t_o at which the step occurs to the time $(t_o + t_s)$, where t_s is the settling time of the transient error response to the step (time for the error transient to fall to **2%** of its initial value). Obviously, the step can be ignored if it is small compared to the remaining terms of the expansion in the interval $t_o < t < (t_o + t_s)$.

Biernson [23] has suggested that the foregoing difficulty can be resolved by examining r, r', r'', . . . for discontinuities and subtracting these discontinuities from the corresponding functions. The remaining functions will all be continuous, and the expansion can be applied over the entire time range of interest. Then, the effects of the discontinuities in r, r', r'', . . . are added to the response. In this procedure, a discontinuity in a function is considered to occur if the function rises (or decays) more abruptly than the corresponding transient response to the discontinuity. In comparing the rise rate of the two curves, a convenient criterion is to compare the times for the two curves to reach **63%** of the initial rise (or decay) of the curves.

A convenient procedure for determining the error coefficients required to satisfy performance specifications is the following:

(a) Given the input $r(t)$ and the maximum allowable error e_{max} which can be tolerated at any time, determine the derivatives r', r'', r''', . . . of the input $r(t)$.

(b) Assume values of the error coefficients so that the maximum value of each component term in the expansion [Eq. **(7-60)**] is equal to e_{max}.

(c) Add the curves obtained in (b) to obtain the first trial value of the error response $e_1(t)$.

(d) There will be times in which the first trial $e_1(t)$ will exceed e_{max}. Referring to the curves found in (b), decide which of the functions r, r', r'' have their maxima in regions where $e_1(t)$ exceeds e_{max}.

(e) Reduce the assumed values of the error coefficients in **(b)** associated with the functions found in (d).

(f) Add the adjusted curves found in (e) to those functions [found in (b)] which have not been changed. Determine whether e_{max} is now exceeded and, if so, repeat (d) and (e).

Once the error coefficients have been specified by means of a procedure similar to the foregoing, the design of the system can be carried out as follows :

(a) The system is designed to meet all other specifications on transient response and frequency response such as bandwidth, M_p, P_{ov}, etc.

(b) The error coefficients of the system designed in (a) are found in terms of the system parameters and the gain.

(c) If any of the error coefficients found in (b) exceed their specified values, they may be reduced by increasing the system gain, if possible, by the introduction of low-frequency dipole ***lead*** functions in the pole-zero configuration of the ***open-loop*** transfer function $G(s)$, or by the feedback-compensation technique suggested by King.[17]

(d) The specifications on transient response and frequency response are rechecked, and step *(c)* is modified if necessary.

A warning should be added. Whenever attempts are made to reduce one or more error coefficients of a system by the methods suggested above, it is possible that higher-order error coefficients may increase. Therefore, if by the addition of a low-frequency lead dipole, an error coefficient can be reduced to zero, a check should be made to insure that higher coefficients have not been increased excessively. In addition, it is generally true that low-frequency poles in a transfer function tend to increase the settling time of the response of the system to steps in the higher input-derivative functions. Therefore, if the actual input being examined has discontinuities in one or more of its higher derivatives, the effect of the longer settling time in the response to these discontinuities must be determined.

7-3 PERFORMANCE INDICES[1,2,9,14,16,18]

A ***performance index P*** is a single number which is used as an ***indirect*** measure of system performance. Other measures of system performance have already been considered, such as the various commonly used parameters M_p, bandwidth, rise time, peak overshoot, etc. However, these parameters provide only a partial description of performance since, in a sense, only part of the corresponding response is described by each. To be sure, if enough of these ***"response parameters"*** (for want of a better term) are known, an accurate description of the corresponding response is possible. That is, the "response parameters" may be considered ***direct*** descriptions of the shape of their associated responses. However, since a response function is continuous, theoretically an infinite number of response parameters are necessary to describe the response. To get around the use of a multitude of response parameters, a performance index may be used. The use of performance indices is an attempt to replace the functional description of the performance of a system through its response parameters by a numerical description that rates the system performance with a single number.

Paragraphs 8-1 and 8-2 describe various techniques for using performance indices. This section merely presents the commonly used indices together with the input conditions for which they apply. Table 7-1 is a summary of these indices. In practice, the performance index corresponding to the specified input is found, and the system is adjusted to optimize (minimize **or** maximize) the index. The indices P_1, P_3, P_5, P_6, P_7, and P_8 can be used in purely analytical procedures. However, P_2 and P_4 are not treated analytically but rather through the use of analog computers.

TABLE 7-1 COMMON PERFORMANCE INDICES

Index	Input	Description or Name	Reference
$P_1 = \int_0^\infty e\,dt$	Transient	Control area	1,2 *
$P_2 = \int_0^\infty \lvert e \rvert\,dt$	Transient	Integral absolute error (IAE)	9
$P_3 = \int_0^\infty t\,e\,dt$	Transient		1
$P_4 = \int_0^\infty t \lvert e \rvert\,dt$	Transient	Integral-time-multiplied absolute error (ITAE)	14,16,18
$P_5 = \int_0^\infty e^2\,dt$	Transient	Integral-square error (ISE)	9,18,33,34
$P_6 = \int_0^\infty te^2\,dt$	Transient		
$P_7 = P_3/P_1^2$	Transient		1
$P_8 = \overline{e^2}$	Stochastic	Mean-square error (MSE)	34,35

* BIBLIOGRAPHY

1 P. T. Nims, "Some Design Criteria for Automatic Controls", *Trans. AIEE,* Vol. 70, Part I, pp. #606-611, 1951.

2 T. M. Stout, "A Note on Control Area", *J. Appl. Phys.,* Vol. 21, pp. #1129-1131, November, 1950.

3 M. R. Aaron, "Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed-Loop Function", *Trans. AIEE,* Vol. 70, Part 11, pp. #1439-1445, 1951.

4 D. Herr and I. Gerst, "The Analysis and an Optimum Synthesis of Linear Servomechanisms", *Trans. AIEE,* Vol. 66, pp. #959-970, 1947.

5 J. H. Mulligan, Jr., "The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems", *Proc. IRE,* Vol. 37, pp. #516-529, May, 1949.

6 D. W. Russel and C. H. Weaver, "Synthesis of Closed-Loop Systems Using Curvilinear Squares to Predict Root Location", *Trans. AIEE,* Vol. 71, Part 11, pp. #95-104, 1952.

7 H. Harris, Jr., M. J. Kirby, and E. F. Von Arx, "Servomechanism Transient Performance from Decibel-Log Frequency Plots", *Trans. AZEE,* Vol. 70, Part 11, pp. #1452-1459, 1951.

8 H. Chestnut and R. W. Mayer, "Comparison of Steady-State and Transient Performance of Servomechanisms", *Trans. AIEE,* Vol. *68,* Part I, pp. #765-777, 1949.

9 F. C. Fickeison and T. M. Stout, "Analogue Methods for Optimum Servomechanism Design", *Trans. AIEE,* Vol. 71, Part 11, pp. #244-250, 1952.

10 M. J. Kirby and D. C. Beaumariage, "Relative Stability of Closed-Loop Systems", *Trans. AZEE,* Vol. *72,* Part 11, pp. #22-43, 1953.

11 G. Biernson, "Quick Methods for Evaluating the Closed-Loop Poles of Feedback Control Systems", *Trans. AIEE,* Vol. 72, Part 11, pp. #53-70, 1953.

12 N. L. Kusters and W. J. M. Moore, "A Generalization of the Frequency Response Method for Study of Feedback Control Systems", in *Automatic and Manual Control,* edited by A. Tustin, Butterworth Scientific Publications, London, England, 1952.

13 Y. Chu, "Correlation Between Frequency and Transient Responses of Feedback Control Systems", *Trans. AIEE,* Vol. 72, Part 11,pp. #81-92, 1953.

14 D. Graham and R. C. Lathrop, "The Synthesis of 'Optimum' Transient Response : Criteria and Standard Forms", *Trans. AZEE,* Vol. 72, Part 11, pp. #273-288, 1953.

15 J. G. Linvill, "The Approximation with Rational Functions of Prescribed Magnitude and Phase Characteristics", Proc. *IRE,* Vol. 40, pp. #711-721, 1952.

16 R. C. Lathrop and D. Graham, "The Transient Performance of Servomechanisms with Derivative and Integral Control", *Trans. AZEE,* Vol. 73, Part 11, pp. #10-17, 1954.

17 L. H. King, "Reduction of Forced Error in Closed-Loop Systems", *Proc. IRE,* Vol. 41, pp. #1037-1042, August, 1953.

18 D. Graham and R. C. Lathrop, "The Influence of Time Scale and Gain on Criteria for Servomechanism Performance", *Trans. AZEE,* Vol. 73, Part 11, pp. #153-158, 1954.

19 A. H. Zemanian, "Bounds Existing on the Time and Frequency Responses of Various Types of Networks", *Proc. IRE,* Vol. 42, pp. #835-839, May, 1954.

20 A. H. Zemanian, "Further Effects of the Pole and Zero Locations on the Step Response of Fixed, Linear Systems", *Trans. AZEE,* Vol. 74, Part 11, pp. #52-55, 1955.

21 H. Thal-Larsen, "Frequency Response from Experimental Nonoscillatory Transient-Response Data", *Trans. AIEE,* Vol. 74, Part 11, pp. #109-113, 1955.

22 D. W. St. Clair, "Step Response as a Short Cut to Frequency Response", *Proc. ZSA,* Vol. 7, pp. #96-101, 1952.

23 G. A. Biernson, "A Simple Method for Calculating the Time Response of a System to an Arbitrary Input", *Trans. AIEE,* Vol. 74, Part 11, pp. #227-245, 1955.

24 M. E. Clynes, "Simple Analytic Method for Linear Feedback System Dynamics", *Trans. AZEE,* Vol. 4, Part 11, pp. #377-382, 1955.

25 G. A. Biernson, "Estimating Transient Response from Open-Loop Frequency Response", *Trans. AZEE,* Vol. 74, Part 11, pp. #388-402, 1955.

26 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 1, pp. #398-439, John Wiley & Sons, Inc., New York, N. Y., 1951.

27 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 2, pp. #1-35, 43-65, John Wiley & Sons, Inc., New York, N.Y., 1955.

28 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #34-87, 278-296, 344-390, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

29 R. C. Seamans, Jr., B. P. Blasingame, and G. C. Clementson, "The Pulse Method for the Determination of Aircraft Performance", J. *Aeronaut. Sci.,* Vol. 17, pp. #22-38, January, 1950.

30 E. A. Guillemin, "Computational Techniques which Simplify the Correlation between Steady-State and Transient Responses of Filters and Other Networks", *Proc. Nat. Electronics Conf.* 1953, Vol. 9, 1954.

31 J. L. Bower, "A Note on the Error Coefficients of a Servomechanism", *J. Appl. Phys.*, Vol. 21, p. #723, July, 1950.

32 E. Arthurs and L. Martin, "A Closed Expansion of the Convolution Integral (A Generalization of Servomechanism Error Coefficients)", *J. Appl. Phys.*, Vol. 26, pp. #58-60, January, 1955.

33 A. C. Hall, *The Analysis and Synthesis of Linear Servomechanisms,* The Technology Press, Massachusetts Institute of Technology, Cambridge, Mass., 1943.

34 G. C. Newton, Jr., "Compensation of Feedback-Control Systems Subject to Saturation", *J. Franklin Inst.*, Vol. 254, pp. #281-296, 391-413, 1952.

35 J. H. Laning, Jr., and R. H. Battin, *Random Processes in Automatic Control,* McGraw-Hill Book Co., Inc., New York, N.Y., 1956.

36 G. S. Brown and D. P. Campbell, *Principles of Servomechanisms,* John Wiley & Sons, Inc., New York, N.Y., 1948.

37 J. G. Linvill, "The Selection of Network Functions to Approximate Prescribed Frequency Characteristics", *MIT Research Laboratory of Electronics Tech Rept. 145,* March 14, 1950.

38 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 1, pp. #401 and 417-434, John Wiley & Sons, Inc., New York, N.Y., 1951.

CHAPTER 8

OPTIMIZATION METHODS FOR TRANSIENT AND STOCHASTIC INPUTS*

8-1 CRITERIA OF PERFORMANCE

In Par. 7-3 it was indicated that the conventional measures of performance such as rise time, peak overshoot, solution time, M_p, etc. were merely partial descriptions of the frequency response of a system or the shape of a particular transient response. As a result, an adequate description of system behavior requires a fair number of response or performance parameters. To avoid using a multiplicity of response parameters (numbers which describe the response such as M_p, peak overshoot, bandwidth, rise time, etc.) attempts have been made to describe system behavior in terms of performance indices. A performance index is a single number which can be used as a criterion of performance. The pertinent performance indices used are those directly related to system error since *error* is the basic determinant of the "goodness" of a system. The most common performance indices are those listed in Par. 7-3.

When performance indices are used in system design, the usual procedure is to minimize the index if it is a direct measure of error. With a given index one also associates the specified input to the system. Several approaches can be used in carrying out the minimization procedure.

In one approach it may be assumed that all but a few of the system parameters are specified. Then, the optimization procedure involves the adjustment of the *free* parameters so as to minimize the performance index. Such a procedure is called a *fixed-configuration minimization* method or technique since the form of the system is specified and only the numerical values of the free parameters are sought.

In another approach nothing is assumed about the configuration of the system. Here the entire impulse response of the system is varied to minimize the performance index. This procedure is called a *free-configuration minimization method or technique.*

Of the two procedures, the easier one to apply and the one more commonly used is the fixed-configuration technique since the process of minimization can be carried out by differentiating the performance index with respect to the free parameters and setting the resulting partial derivatives equal to zero. The fixed-configuration technique is also easy to apply when use is made of an analog computer. The free-configuration method, on the other hand, is less commonly used because it can only be applied by the use of the calculus of variations since in this case a system *function* (rather than a system parameter) is varied to obtain a minimum.

In practice, the application of optimization methods can lead to failure when one is not cognizant of the limitations of the mathematical model that represents the physical system. The optimum system often requires cancellation of the characteristics of the fixed elements of the system, resulting in an unnecessarily wide-band performance and concurrent nonlinear operation. To avoid this, constraints may be placed on signal levels or on

By L.A. Gould

bandwidth. Constrained optimization brings the designer closer to the practical limitations of the system and serves to guide system design in a realistic way.

The primary advantages of optimization procedures (as contrasted with conventional trial-and-error procedures) are twofold. First, the designer is able, through optimization, to decide whether a given set of specifications is compatible. Second, the designer can decide whether a compatible set of specifications can be satisfied when bounded by constraints.

8-2 OPTIMUM SYNTHESIS OF FIXED-CONFIGURATION SYSTEMS

8-2.1 TRANSIENT INPUTS

For transient inputs the integral-square error criterion (ISE) is commonly used[5,7,10] to obtain optimum synthesis of a fixed-configuration system. If discussion is limited to unity-feedback systems, where the desired output is the input, then the actuating signal $e(t)$ is equal to the system error $y_e(t)$. The ISE criterion is then

$$I_y = \int_{-\infty}^{+\infty} y_e^2(t)\, dt \tag{8-1}$$

The evaluation of the integral in Eq. (8-1) is facilitated by the application of Parseval's theorem :

$$\int_{-\infty}^{+\infty} x^2(t)\, dt = \frac{1}{2\pi j}\int_{-\infty}^{+\infty} X(s)X(-s)\, ds \tag{8-2}$$

where $X(s)$ is the Fourier transform of $x(t)$. Thus, Eq. (8-1) can be written

$$I_y = \frac{1}{2\pi j}\int_{-\infty}^{+\infty} Y_e(s)Y_e(-s)\, ds \tag{8-3}$$

The procedure for minimizing the ISE is as follows :

(a) Express the Fourier transform of the error as a function of the complex frequency s. This function will involve the free parameters of the system as unknown coefficients.

(b) Express I_y in terms of $Y_e(s)$ by means of Eq. (8-3). If $Y_e(s)$ is rational, the form of I_y will be

$$I_y = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} \frac{C(s)\, ds}{D(s)D(-s)} \tag{8-4}$$

where $C(s)$ and $D(s)$ are polynomials in s.

(c) Evaluate the integral in Eq. (8-4). Definite integrals of this form have been evaluated in terms of the coefficients of the polynomials in the integrand.[8,15] A brief table of such integrals is presented in Table 8-1 where the evaluation has been carried out for $s = j\omega$. At this point, I_y is expressed as a function of the free parameters p_1 through p_k, i.e.,

$$I_y = I_y(p_1, p_2, \ldots, p_k) \tag{8-5}$$

(d) Adjust the free parameters $p_1, p_2, \ldots$ so as to minimize I_y. This can be accomplished analytically by solving the k simultaneous equations

$$\frac{\partial I_y}{\partial p_i} = 0 \quad (i = 1, 2, \ldots, k) \tag{8-6}$$

However, it is often better to find the minimum graphically by working directly with I_y.

Example. A unity-feedback system has the fixed-element transfer function

$$G_f(s) = \frac{1}{s(T_f s + 1)(T_m s + 1)}$$

where

$T_f = 0.01$ second and

$T_m = 0.04$ second

The compensation $G_c(s)$ is a pure gain, i.e.,

$G_c(s) = K_v$, the velocity constant

The input is a step of magnitude N_i, i.e.,

$r(t) = N_{i\delta-1}(t)$

The desired output is the input. The configuration that describes the problem appears in Fig. 8-1.

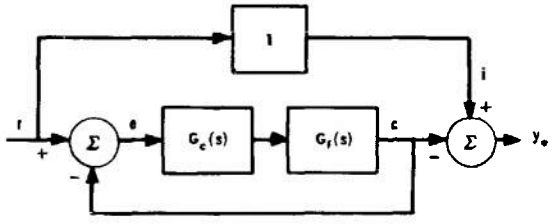


Fig. 8-1 Configuration for ISE minimization.

Solution. To find the value of K, that minimizes the integral-square error I_v, we first find the error transform $Y_e(s)$. From the configuration in Fig. 8-1 it is evident that

$$Y_e(s) = E(s) = R(s) - C(s), \text{ and}$$

$$C(s) = \frac{G_c(s)G_f(s)}{1 + G_c(s)G_f(s)} R(s)$$

Now, $R(s) = \dfrac{N_i}{s}$.

So, by substituting the expressions given originally for $G_c(s)$ and $G_f(s)$ into the equation for $C(s)$ we find that the error transform is

$$Y_e(s) = N_i \frac{T_f T_m s^2 + (T_f + T_m)\, s + 1}{T_f T_m s^3 + (T_f + T_m) s^2 + s + K_v}$$

Substituting $Y_e(s)$ into Eq. (8-3) and letting $s = j\omega$, we find that

where

$c_0 = 1$ $\qquad$ $d_0 = K_v$

$c_2 = T_f^2 + T_m^2$ $\qquad$ $d_1 = 1$

$c_4 = T_f^2 T_m^2$ $\qquad$ $d_2 = T_f + T_m$

$\qquad\qquad\qquad\quad$ $d3 = T_f T_m$

Using I_3 of Table 8-1 to evaluate the integral above, we get

$$I_v = \frac{N_i^2}{2} \left\{ \frac{1 + \dfrac{T_f}{T_m} + \dfrac{T_m}{T_f} + \dfrac{1}{K_v}\left(\dfrac{1}{T_f} + \dfrac{1}{T_m}\right)}{\dfrac{1}{T_m} + \dfrac{1}{T_f} - K_v} \right\}$$

Numerically, this becomes

$$I_v = \frac{N_i^2}{2} \left[\frac{5.25K_v + 125}{125K_v - K_v^2} \right]$$

Inspection shows that $I_v \to 0$ if $K_v \to \infty$, but this solution is not allowed since the system would then be unstable. Differentiating I_v with respect to K_v and setting $dI_v/dK_v = 0$ yields

$$K_v^2 + 47.6\, K_v - 2980 = 0$$

or

$$K_v = 35.8 \text{ or } -83.4$$

The negative value is not allowed so a velocity constant $K_v = 35.8$ sec^{-1} minimizes the integral-square error. The value of the minimum integral-square error is

$$I_{v_{min}} = 0.049\, N_i^2$$

As a point of interest, for $K_v = 35.8$ sec^{-1}, the value of the peak magnification is 10 $\log_{10} M_p = 2.8$ dg which is a reasonable value.

Another optimization criterion is presented in a series of papers by Graham and Lathrop[3,4,5] in which they have applied the integral-time-multiplied-absolute-error criterion (ITAE) to optimize the performance of

$$I_v = \frac{N_i^2}{2\pi} \int_{-\infty}^{+\infty} \frac{[c_4\omega^4 + c_2\omega^2 + c_0]\, d\omega}{[d_3(j\omega)^3 + d_2(j\omega)^2 + d_1(j\omega) + d_0]\,[d_3(-j\omega)^3 + d_2(-j\omega)^2 + d_1(-j\omega) + d_0]}$$

standard-system forms. However, their procedure is limited to step inputs only. The ITAE criterion is

$$I_{ta} = \int_{-\infty}^{+\infty} t|y_e(t)|\,dt \qquad (8\text{-}7)$$

Although the analytical application of the ITAE criterion is practically impossible, the performance index can be easily mechanized on an analog computer.

For systems exhibiting zero steady-state error for a step input (finite velocity constant), the standard form chosen was

$$\frac{C(s)}{R(s)} = \frac{1}{s^n + q_{n-1}s^{n-1} + \ldots + q_1 s + 1} \qquad (8\text{-}8)$$

The denominator polynomials of the optimum systems (that minimize the ITAE for a step input) are listed in Table 8-2 for the first eight orders. Figure 8-2 shows the step responses of the optimum systems, and Fig. 8-3 shows the frequency responses of these systems.

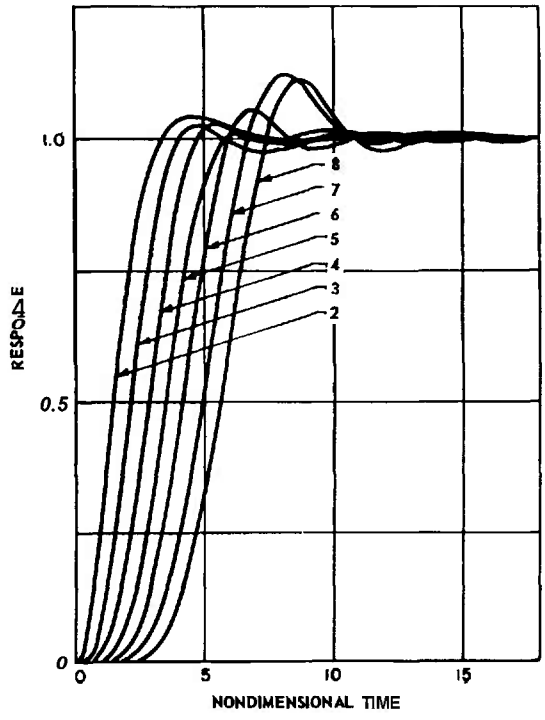


Fig. 8-2 Step-function responses of the optimum unit-numerator transfer systems, second to eighth orders.

By permission from *Transactions* of the *AIEE*. Volume 72, Part II, 1953, from article entitled 'The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms', by Dunstan Graham and R. C. Lathrop.

TABLE 8-1 TABLE OF DEFINITE INTEGRALS

$$I_n = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{C_{2n-2(\omega)}}{D_n(j\omega)\,D_n(-j\omega)}\,d\omega$$

where

$$C_{2n-2(\omega)} = c_{2n-2}\omega^{2n-2} + c_{2n-4}\omega^{2n-4} + \ldots + c_2\omega^2 + c_0$$

$$D_n(j\omega) = d_n(j\omega)^n + d_{n-1}(j\omega)^{n-1} + \ldots + d_1 j\omega + d_0$$

$$I_1 = \frac{c_0}{2d_1d_0}$$

$$I_2 = \frac{c_2d_0 + d_2c_0}{2d_2d_1d_0}$$

$$I_3 = \frac{c_4d_1d_0 + c_2d_3d_0 + c_0d_3d_2}{2d_3d_0(d_2d_1 - d_3d_0)}$$

$$I_4 = \frac{(d_2d_1 - d_3d_0)\,d_0c_6 + d_4d_1d_0c_4 + d_4d_3d_0c_2 + (d_3d_2 - d_4d_1)\,d_4c_0}{2d_4d_0(d_3d_2d_1 - d_4d_1^2 - d_3^2d_0}$$

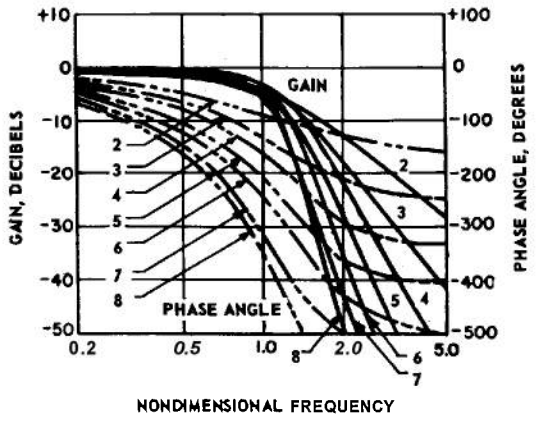


Fig. 8-3 Frequency responses of the optimum unit-numerator transfer systems.

By permission from *Transactions of the AZEE,* Volume *72,* Part II, 1953, from article entitled 'The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms', by Dunstan Graham and R. C. Lathrop.

For systems exhibiting zero steady-state error for a ramp input (infinite velocity constant, finite acceleration constant), the standard form chosen was

$$\frac{C(s)}{R(s)} = \frac{q_1 s + 1}{s^n + q_{n-1}s^{n-1} + \ldots + q_2 s^2 + q_1 s + 1} \tag{8-9}$$

The denominator polynomials of the optimum systems of this type are listed in Table 8-3 for the second to sixth orders. The step responses of the optimum systems are shown in Fig. 8-4.

Additional matter such as optimum compensation of various systems and the effect of time scaling are also discussed by these authors[4,5] with the use of the ITAE criterion as the performance index.

TABLE 8-2 THE MINIMUM ITAE STANDARD FORMS, ZERO-DISPLACEMENT-ERROR SYSTEMS

$$s + \omega_0$$
$$s^2 + 1.4\omega_0 s + \omega_0^2$$
$$s^3 + 1.75\omega_0 s^2 + 2.15\omega_0^2 s + \omega_0^3$$
$$s^4 + 2.1\omega_0 s^3 + 3.4\omega_0^2 s^2 + 2.7\omega_0^3 s + \omega_0^4$$
$$s^5 + 2.8\omega_0 s^4 + 5.0\omega_0^2 s^3 + 5.5\omega_0^3 s^2 + 3.4\omega_0^4 s + \omega_0^5$$
$$s^6 + 3.25\omega_0 s^5 + 6.60\omega_0^2 s^4 + 8.60\omega_0^3 s^3 + 7.45\omega_0^4 s^2 + 3.95\omega_0^5 s + \omega_0^6$$
$$s^7 + 4.475\omega_0 s^6 + 10.42\omega_0^2 s^5 + 15.08\omega_0^3 s^4 + 15.54\omega_0^4 s^3 + 10.64\omega_0^5 s^2 + 4.58\omega_0^6 s + \omega_0^7$$
$$s^8 + 5.20\omega_0 s^7 + 12.80\omega_0^2 s^6 + 21.60\omega_0^3 s^5 + 25.75\omega_0^4 s^4 + 22.20\omega_0^5 s^3 + 13.30\omega_0^6 s^2 + 5.15\omega_0^7 s + \omega_0^8$$

By permission from *Transactions of the AZEE.* Volume *72,* Part II, 1953, from article entitled 'The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms', by Dunstan Graham and R. C. Lathrop.

TABLE 8-3 THE MINIMUM ITAE STANDARD FORMS, ZERO-VELOCITY-ERROR SYSTEMS

$$s^2 + 3.2\omega_0 s + \omega_0^2$$
$$s^3 + 1.75\omega_0 s^2 + 3.25\omega_0^2 s + \omega_0^3$$
$$s^4 + 2.41\omega_0 s^3 + 4.93\omega_0^2 s^2 + 5.14\omega_0^3 s + \omega_0^4$$
$$s^5 + 2.19\omega_0 s^4 + 6.50\omega_0^2 s^3 + 6.30\omega_0^3 s^2 + 5.24\omega_0^4 s + \omega_0^5$$
$$s^6 + 6.12\omega_0 s^5 + 13.42\omega_0^2 s^4 + 17.16\omega_0^3 s^3 + 14.14\omega_0^4 s^2 + 6.76\omega_0^5 s + \omega_0^6$$

By permission from *Transactions of the AZEE* Volume *72,* Part II, 1953, from article entitled 'The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms', by Dunstan Graham and R. C. Lathrop.

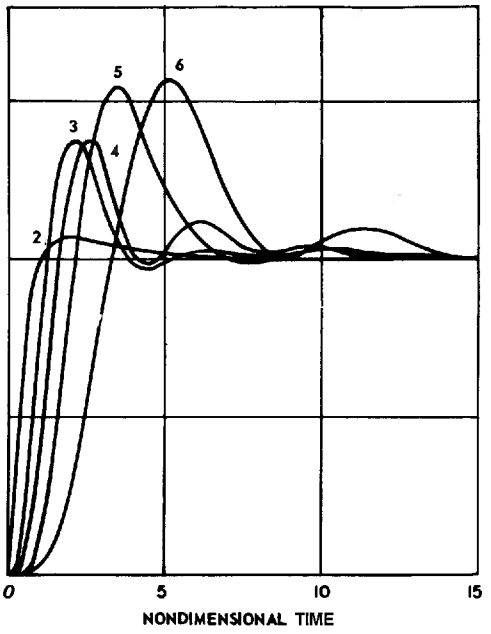


Fig. 8-4 Step-function responses of the optimum zero-velocity-error systems, second to sixth orders.

By permission from *Transactions of the AZEE.* Volume 72, Part II, 1963, from article entitled 'The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms', by Dunstan Graham and R. C. Lathrop.

8-2.2 STATIONARY STOCHASTIC INPUTS[8,12,15]

The mean-square error (MSE) criterion is universally used as a performance index when the input is stochastic. The general configuration that applies to this minimization problem is shown in Fig. 8-5.

In this figure, v *(t)* is the data component of the input, and *n(t)* is the noise component of the input. The mean-square error is defined as

$$\overline{y_e^2} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} y_e^2(t)\, dt \qquad (8\text{-}10)$$

On the assumption that the data and the noise are uncorrelated, application of the formulae given in Par. 3-8 to the configuration of Fig. 8-5 yields

$$\overline{y_e^2} = \int_{-\infty}^{+\infty} \Phi_{yy}(\omega)\, d\omega \qquad (8\text{-}11)$$

where

$$\Phi_{yy}(s) = [1 - W(s)]\,[1 - W(-s)]\,\Phi_{vv}(s) + W(s)\,W(-s)\,\Phi_{nn}(s) \qquad (8\text{-}12)$$

$$W(s) = \frac{G_c(s)\,G_f(s)}{1 + G_c(s)\,G_f(s)} \qquad (8\text{-}13)$$

$\Phi_{yy}(s)$ = power spectrum of system error y_e

$\Phi_{vv}(s)$ = power spectrum of data v

$\Phi_{nn}(s)$ = power spectrum of noise *n*

Since $\Phi_{yy}(\omega)$ is an even function, the evaluation of the integral in Eq. (8-11) can be carried out by means of the integral table (Table 8-1).

In all other respects, the design procedure for minimizing the mean-square error for stationary stochastic inputs with a fixed system configuration parallels the procedure for transient inputs outlined above.

Example. For the configuration of Fig. 8-5,

$$G_f(s) = \frac{1}{s(T_f s + 1)(T_m s + 1)}$$

$$G_c(s) = K_v$$

$$\Phi_{vn}(s) = 0$$

$$\Phi_{vv}(s) = \frac{\sigma_v^2 \upsilon/\pi}{-s^2(\upsilon^2 - s^2)}$$

$$\Phi_{nn}(s) = \frac{\gamma_n}{\pi} \text{ (white noise)}$$

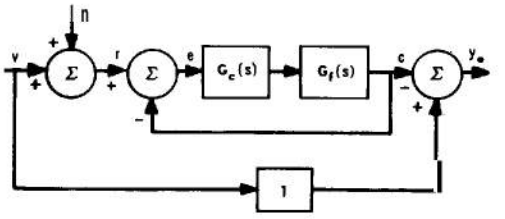


Fig. 8-5 Configuration for MSE minimization.

$T_f = 0.01$ sec

$T_m = 0.04$ sec

$\sigma_v = 10$ milliradian/sec

$v = 0.1\ (\text{sec})^{-1}$

$\gamma_n = 0.4\ (\text{milliradian})^2 - \text{sec}$

If we assume that there is no noise initially ($\Phi_{nn} = 0$), then

$$1 - W(s) = \frac{s(T_f s + 1)(T_m s + 1)}{T_f T_m s^3 + (T_f + T_m) s^2 + s + K_v}$$

A normalized frequency is chosen such that, for $s = j\omega$, $T_f \omega = u$. By applying the normalization theorem of Par. 3-4 (Eq. 3-46) to the expression for $1 - W(s)$ above, we obtain

$$1 - W(ju) = \frac{(1 + ju)(1 + bju)}{ju\, b(ju)^3 + (1 + b)(ju)^2 + (ju) + K}$$

where

$$b = \frac{T_m}{T_f} = 4$$

$$K = T_f K_v$$

In addition, the power spectrum of data expressed in terms of u is found to be

$$\Phi_{vv}(u) = \left[\frac{\sigma_v^2 T_f^3}{\pi}\right]\left[\frac{a}{u^2(a^2 + u^2)}\right]$$

where

$$a = T_f v = 0.001$$

From Eq. (8-12), the power spectrum of system error expressed in terms of u is found to be

$$\Phi_{yy}(u) = \left[\frac{a\sigma_v^2 T_f^3}{\pi}\right]\left[\frac{c(u)}{D_4(ju) D_4(-ju)}\right]$$

where

$$C(u) = b^2 u^4 + (1 + b^2) u^2 + 1$$

$$D_4(ju) = b(ju)^4 + (1 + b + ab)(ju)^3 + (1 + a + ab)(ju)^2 + (a + K)(ju) + aK$$

$$D_4(-ju) = b(-ju)^4 + (1 + b + ab)(-ju)^3 + (1 + a + ab)(-ju)^2 + (a + K)(-ju) + aK$$

Then, using Eq. (8-11), we find the mean square error to be

$$\overline{y_e^2} = 2a\sigma_v^2 T_f^2 \frac{1}{2\pi}\left\{\int_{-j\infty}^{+\infty} \frac{c(u)}{D_4(ju) D_4(-ju)}\, du\right\}$$

Evaluating this expression from I_4 of the integral tables (Table 8-1) and substituting numerical values, we find that

$$\left[\frac{\sqrt{\overline{y_e^2}}}{\sigma_v T_f}\right]^2 \triangleq N = \frac{0.016K^2 - 3.99K + 5.015}{-4K^3 + 5K^2 + 0.005K}$$

To determine the minimum value of N it is convenient to make a plot of N versus K. This avoids a differentiation of N with respect to K which results in a fourth-degree algebraic equation whose roots must then be determined. It is evident that using the plot to determine the minimum is a simpler technique. The minimum from the plot is found to occur at $K = 1.1$ or $N = 0.90$. Consequently, the optimum system has

$K_v = 90\ \text{sec}^{-1}$

$$\left[\overline{y_e^2}\right]^{1/2} = 0.095 \text{ milliradian}$$

If the noise is considered, the procedure is more involved but unchanged in principle. The results of the minimization of $\overline{y_e^2}$ with the noise added to the data as follows:

$K_v = 7.8\ \text{sec}^{-1}$

$$\left[\overline{y_e^2}\right]^{1/2} = 2.21 \text{ milliradian}$$

8-3 OPTIMUM SYNTHESIS OF FREE-CONFIGURATION SYSTEMS WITH STATIONARY STOCHASTIC INPUTS[6,7,11,12,13,15]

The design problem for the optimum synthesis of free-configuration systems with stationary stochastic inputs is one of determining the closed-loop transfer function

$$\frac{C(s)}{R(s)} = W(s) \tag{8-14}$$

that minimizes the mean-square error when information is given concerning the data $v(t)$, noise $n(t)$, desired output $i(t)$, and fixed-element transfer function $G_f(s)$. No information concerning the form of the compensation $G_c(s)$ is needed. Figure 8-6 shows the configuration that describes the problem (G_d is the ideal element transfer function).

The solution to this problem is obtained by means of the calculus of variations and is in the form of an integral equation:

$$\int_{-\infty}^{\infty} dt_3 g_f(t_3) \int_{-\infty}^{\infty} dt_4 g_f(t_4) \int_{-\infty}^{\infty} dt_2 \omega(t_2)\, \phi_{rr}(t_1 + t_3 - t_2 - t_4) - \int_{-\infty}^{\infty} dt_3 g_f(t_3)\, \phi_{ri}(t_1 + t_3) = 0 \text{ for } t_1 \geqq 0 \tag{8-15}$$

where

$g_f(t)$ = impulse response of fixed elements

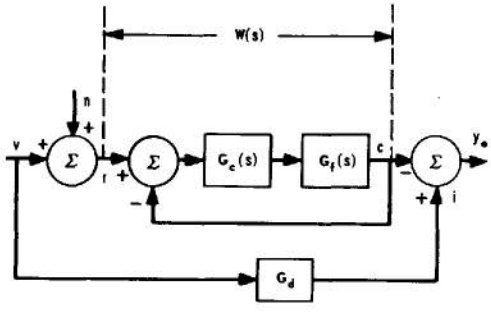


Fig. 8-6 Configuration for MSE minimization

$w(t)$ = impulse response of control system ($\mathcal{L}^{-1}[W(s)]$)

$\phi_{rr}(t)$ = autocorrelation function of $r(t)$

$\phi_{ri}(t)$ = crosscorrelation function between $r(t)$ and $i(t)$

If the fixed elements are minimum phase (no zeros in right-half s-plane) or are unspecified, Eq. (8-15) reduces to the Wiener-Hopf equation:[6,12,15]

$$\int_{-\infty}^{\infty} dt_2 \omega(t_2)\, \phi_{rr}(t_1 - t_2) - \phi_{ri}(t_1) = 0 \text{ for } t_1 \geqq 0 \tag{8-16}$$

If the autocorrelation functions and the impulse response of the fixed elements are Fourier transformable, Eqs. (8-15) and (8-16) can be solved in terms of transforms by a method called spectrum factorization. For Eq. (8-15), the optimum system function $W(s)$ is given by

$$W(s) = \frac{\left[\dfrac{\Gamma(s)}{\Delta^-(s)}\right]_+}{\Delta^+(s)} \tag{8-17}$$

where

$\Gamma(s) = 2\pi G_f(-s)\, \Phi_{ri}(s)$

$\Delta(s) = 2\pi G_f(s)\, G_f(-s)\, \Phi_{rr}(s)$

$\Delta^+(s) \triangleq$ that factor of $\Delta(s)$ which contains all the poles and zeros of $\Delta(s)$ which lie in the left-hand s-plane

$\Delta^-(s) = \dfrac{\Delta(s)}{\Delta^+(s)}$

$\left[\dfrac{\Gamma(s)}{\Delta^-(s)}\right]_+ \triangleq$ that part of the *partial-fraction expansion* of $\Gamma(s)/\Delta^-(s)$ due to the poles of $\Gamma(s)/\Delta^-(s)$ which lie in the left-half s-plane

$$\left[\frac{\Gamma(s)}{\Delta^-(s)}\right]_- = \frac{\Gamma(s)}{\Delta^-(s)} - \left[\frac{\Gamma(s)}{\Delta^-(s)}\right]_+$$

Transformation and factorization of Eq. (8-16) yields

$$W(s) = \frac{\left[\frac{\Phi_{ri}(s)}{\Phi^{-}_{rr}(s)}\right]}{\Phi^{+}_{rr}(s)} \qquad (8\text{-}18)$$

where the notation is the same as that defined below Eq. (8-17).

Example. $G_f(s)$ is minimum phase, and

$$\Phi_{vv}(s) = \frac{\sigma_v^2 \nu}{\mathrm{x}\,(-s^2)\,(\nu^2 - s^2)}$$

$$\Phi_{nn}(s) = \frac{\gamma_n}{\pi}$$

$$\Phi_{vn}(s) = 0$$

$$\sigma_v = 10 \text{ milliradian/sec}$$

$$\nu = 0.1 \text{ sec}^{-1}$$

$$\gamma_n = 0.4 \text{ (milliradian)}^2 - \text{sec}$$

$$i(t) = v(t)$$

Normalize the frequency scale by letting

$$\lambda = \frac{s}{\nu}.$$

Then,

$$\Phi_{vv}(\lambda) = \frac{\sigma_v^2}{\pi\nu^2} \frac{1}{(-\lambda^2)\,(1-\lambda^2)}$$

$$\Phi_{nn}(\lambda) = \frac{\nu\gamma_n}{\pi}$$

Let

$$\frac{\sigma_v^2}{\nu^2} = a \quad \text{and} \quad \nu\,\gamma_n = b$$

Since the data and noise are un orrelat d $(\Phi_{vn} = 0)$,

$$\Phi_{rr}(\lambda) = \Phi_{vv}(\lambda) + \Phi_{nn}(\lambda) = \frac{b}{\pi}\left[\frac{c^2 - \lambda^2 + \lambda^4}{(-\lambda^2)\,(1-\lambda^2)}\right]$$

where

$$c^2 = \frac{a}{b} = \frac{\sigma_v^2}{\nu^3\,\gamma_n} = 25 \times 104$$

Since the desired output $i(t)$ is the data v (t),

$$\Phi_{ri}(\lambda) = \Phi_{vv}(\lambda)$$

Equation (8-18) applies to this problem. To find $\Phi^{-}_{rr}(\mathrm{A})$ and $\Phi^{+}_{rr}(h)$ it is necessary to distinguish between poles and zeros in the two half planes. Since $\Phi_{rr}(\mathrm{A})$ has a double pole on the imaginary axis at the origin, the following artifice is used. We let

$$-\lambda^2 = \lim_{\varepsilon \to 0} (\mathrm{E} - \mathrm{A})\,(\mathrm{E} + \mathrm{A})$$

The problem can then be worked with (E − A) (E + A) replacing $-\lambda^2$. After carrying out the pertinent algebraic manipulations, we let $\varepsilon \to 0$. This is equivalent to factoring $-\lambda^2$ into (− A) (+A) and then associating $(-\lambda)$ with the right-half plane (RHP) and $(+\lambda)$ with the left-half plane (LHP). Therefore, the factorization of @ (A) becomes

$$\Phi_{rr}(\lambda) = \left[\frac{(m + jn + \lambda)\,(m - jn + \lambda)}{(+\lambda)\,(1 + \lambda)}\right] \left[\frac{b}{\pi} \frac{(m + jn - \lambda)\,(m - jn - \lambda)}{(-\lambda)\,(1 - \lambda)}\right]$$

where

$$m = 0.5\sqrt{2c + 1} = 15.82$$

$$n = 0.5\sqrt{2c - 1} = 15.80$$

The factor of $\Phi_{rr}(\lambda)$ having all its poles and zeros in the right-half plane is

$$\Phi^{-}_{rr}(\lambda) = \frac{b}{\pi} \frac{(m + jn - \lambda)\,(m - jn - \lambda)}{(-\lambda)\,(1 - \lambda)}$$

Therefore,

$$\frac{\Phi_{ri}(s)}{\Phi^{-}_{rr}(s)} = \frac{c^2}{(+\lambda)\,(1 + \lambda)\,(m + jn - \lambda)\,(m - jn - \lambda)}$$

This function has left-half-plane poles at A = 0 and $\lambda = -1$. Expanding in terms of partial fractions and retaining only those terms in the expansion due to LHP poles, we obtain

$$\left[\frac{\Phi_{ri}(\lambda)}{\Phi_{rr}^{-}(\lambda)}\right]_{+} = c\left\{\frac{1}{\lambda} - \frac{c}{1+c+\sqrt{2c+1}}\right\}$$

or

$$\left[\frac{\Phi_{ri}(\lambda)}{\Phi_{rr}^{-}(\lambda)}\right]_{+} = (\sqrt{2c+1}-1)\frac{\left(\frac{1+\sqrt{2c+1}}{2}+\lambda\right)}{+\lambda(1+\lambda)}$$

Since,

$$\Phi_{rr}^{+}(\lambda) = \frac{(m+jn+\lambda)(m-jn+\lambda)}{(+\lambda)(1+\lambda)}$$

$$W(\lambda) = (\sqrt{2c+1-1})\frac{\left(\frac{1+\sqrt{2c+1}}{2}+\lambda\right)}{(\lambda^2+\sqrt{2c+1}\,\lambda+c)}$$

Numerically, since A = 10s and c = 500, the optimum response is

$$W(s) = \frac{1+0.613s}{\left(\frac{s}{\omega_n}\right)^2 + 2\zeta\left(\frac{s}{\omega_n}\right)+1}$$

where

$$\zeta = \sqrt{\frac{1}{2}+\frac{1}{4c}} = 0.704$$

$$\omega_n = v\sqrt{c} = 2.24$$

The open-loop transfer function corresponding to the optimum response is

$$\frac{C(\lambda)}{E(\lambda)} = G_c(\lambda)G_f(\lambda) = c\frac{\left(1+\frac{2}{1+\sqrt{2c+1}}\lambda\right)}{\lambda(1+\lambda)}$$

or, numerically,

$$\frac{C(s)}{E(s)} = 50\,\frac{(0.613s+1)}{s(10s+1)}$$

The mean-square error $\overline{y_e^2}$ can then be evaluated by using Eqs. (8-11) and (8-12) together with Table 8-1. The rms error due to noise alone is found to be

$$\left[\overline{y_e^2}\right]_n^{1/2} = 1.35 \text{ milliradian}$$

The rms error due to the signal acting alone is

$$\left[\overline{y_e^2}\right]_v^{1/2} = 0.796 \text{ milliradian}$$

Consequently, the total rms error is

$$\left[\overline{y_e^2}\right]^{1/2} = \sqrt{\left[\overline{y_e^2}\right]_n + \left[\overline{y_e^2}\right]_v}$$

or

$$\left[\overline{y_e^2}\right]^{1/2} = 1.57 \text{ milliradian}$$

8-4 LIMITATIONS AND APPLICATION PROBLEMS

Several difficulties confront the designer who carries out the optimization procedure in any practical problem. He finds that (1) the labor involved becomes excessive, and (2) the resulting optimum compensation is both difficult to realize and unrealistic. The latter difficulty arises because cancellation of the fixed-component transfer function is required, resulting in component saturation and poor utilization of hardware. One factor acting in favor of the designer using the ISE and MSE criteria is that the minima resulting from the use of these procedures are broad. Thus, a fairly wide deviation of parameters and configurations can occur without appreciably altering the performance index. Hence, the labor involved in designing by optimization techniques can be reduced greatly by judicious engineering approximations. In addition, more freedom is available to the designer when the minima that arise in an optimization problem are broad since he can then satisfy additional performance specifications such as rise time, peak overshoot, etc. The same cannot be said for the ITAE criterion since it is a selective criterion producing narrow minima that leave little freedom to the designer. Therefore, the ITAE criterion is not to be recommended if parameter variation is to be expected and other performance specifications are to be met.

Techniques for overcoming the second limitation of optimization procedures have been proposed by Newton.[7,10] He recommends that constraints be placed on the signals that are not to saturate. That is, the optimization is to be carried out by requiring that the performance index be minimized while the signals that may saturate are kept below a specified upper limit. Actually, however, a measure of the peak-signal values is used to facilitate analysis. In the case of transient inputs, the integral-square signal values are to be kept below assigned limits. In the case of stochastic signals, the rms signal values are to be constrained. It is also possible to combine the two types of signals by requiring, for example, that the rms error for a stochastic input be minimized subject to a constraint on the integral-square value of a specified signal for a transient input.

Newton[10] also proposes that constrained optimization be carried out by minimizing bandwidth since a minimum bandwidth system is highly desirable in any case. Thus, bandwidth is minimized subject to a constraint on the allowable error index.

By employing constrained optimization using performance indices that exhibit broad minima, the designer can approach a problem with a greater degree of certainty of finding out whether his specifications are compatible and, if compatible, whether they can be met in practice.

The optimization procedures discussed here have been limited to transient inputs and stationary stochastic inputs. If the input is nonstationary, then the optimum system will become nonstationary or time-variable. If the form of the time variation in the input statistics is known, it is possible to design a system which exhibits variable bandwidth. Unfortunately, there is as yet no general method for finding an explicit analytical solution to this problem. If the time variation in the input statistics is slow compared with the response time of the system, then the nonstationary problem can be broken down into a series of stationary segments. If the input statistics vary at a rate that is of the order of the response time of the system, then one cannot ignore the nonstationary nature of the problem.

BIBLIOGRAPHY

1 P. T. Nims, "Some Design Criteria for Automatic Controls", *Trans. AIEE,* Vol. *70,* Part I, pp. #606-611, 1951.

2 F. C. Fickeison and T. M. Stout, "Analogue Methods for Optimum Servomechanism Design", *Trans. AIEE,* Vol. 71, Part 11, pp. #244-250, 1952.

3 D. Graham and R. C. Lathrop, "The Synthesis of 'Optimum' Transient Response : Criteria and Standard Forms", *Trans. AIEE,* Vol. *72,* Part 11, pp. #273-288, 1953.

4 R. C. Lathrop and D. Graham, "The Transient Performance of Servomechanisms with Derivative and Integral Control", *Trans. AIEE,* Vol. 73, Part 11, pp. #10-17, 1954.

5 D. Graham and R. C. Lathrop, "The Influence of Time Scale and Gain on Criteria for Servomechanism Performance", *Trans. AIEE,* Vol. 73, Part 11, pp. #153-158, 1954.

6 N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series,* John Wiley & Sons, Inc., New York, N.Y., 1949.

7 G. C. Newton, Jr., "Compensation of Feedback Control Systems Subject to Saturation", J. *Franklin Inst.,* Vol. 254, pp. #281-296, 391-413, October, November, 1952.

8 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms,* MIT Radiation Laboratory Series, Vol. 25, pp. #262-370, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

9 J. H. Westcott, "The Introduction of Constraints into Feedback System Designs", *Trans. IRE,* CT-1, pp. #39-49, September, 1954.

10 G. C. Newton, Jr., "Design of Control Systems for Minimum Bandwidth", *Trans. AIEE,* Vol. 74, Part 11, pp. #161-168, 1955.

11 H. S. Tsien, *Engineering Cybernetics,* pp. #111-135, 231-252, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

12 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #410-499, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

13 H. W. Bode and C. E. Shannon, "A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory", *Proc. IRE,* Vol. 38, pp. #417-425, April, 1950.

14 L. A. Zadeh and J. R. Ragazzini, "An Extension of Wiener's Theory of Prediction", J. *Appl. Phys.,* Vol. 21, pp. #645-655, July, 1950.

15 J. H. Laning, Jr. and R. H. Battin, *Random Processes in Automatic Control,* McGraw-Hill Book Company, Inc., New York, N.Y., 1956.

CHAPTER 9

SAMPLED-DATA SYSTEMS*

9-1 GENERAL THEORY

Linvill states, "A sampled-data control system is one wherein the signal supplied to one or more parts of the system is not given continuously in time, but is supplied at discrete values of the time variable, t. In such a system, the part of the system being fed intermittently might, for example, have an input signal applied to it at $t = 0, T, 2T, 3T, \ldots$ (where T is the length of time between samplings) with no data at all supplied in the intervals separating these sampling instants. A control system makes use of sampled data when it is impossible to supply continuous data to all its parts."†

†Quoted by permission from *Transactions of the AZEE,* Volume *70,* Part 11, 1951, from article entitled "Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation," by W. K. Linvill.

If the sampling frequency is high compared to the signal frequency and the critical frequencies of the system, then the fact that the data are sampled has little bearing on system behavior. Otherwise, the effect of sampling may become quite pronounced.

Figure 9-1 shows the elements of a typical sampled-data system. The input $r(t)$ may be composed of sampled or continuous data. The sampling device periodically samples the actuating signal $e(t)$ under control of the carrier signal supplied to it. The holding circuit is used to smooth the sampled output from the sampling device, and the smoothed output of the holding circuit then drives the output member. It is evident that the components and signals in the system are combinations of discrete and continuous elements. Because part

By L. A. Gould

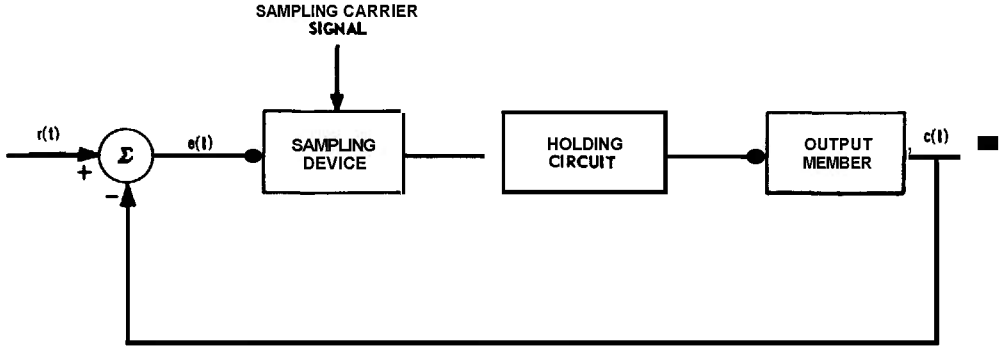


Fig. 9-1 Sampled-data system.

of the system operates on sampled data and part on continuous data, the analysis of system behavior is not easily carried out by conventional methods. For that part of the system operating on continuous data, conventional methods of analysis are best. For that part operating on sampled data, the use of sequences and linear difference equations is best. However, methods have been developed which treat sampled-data systems from a unified viewpoint and will be presented in this chapter.

A sampled-data system like that in Fig. **9-1** can be represented by the mathematical model shown in Fig. 9-2. The impulse modulator is an ideal device that multiplies the actuating signal $e(t)$ by the carrier signal $\Delta(t)$. The function $\Delta(t)$ represents a periodic train of unit impulses occurring at a frequency $\Omega = 2\pi/T$ radians per second (Fig. 9-3)—where Ω is called the sampling frequency. The equivalent linear filter is so chosen that the combined operation of the impulse modulator and equivalent linear filter on the actuating signal in the model produces the same input to the output member as the combined action of the sampling device and holding circuit in the original system. For example, a system in which the actuating signal is sampled every T seconds by a device which holds a particular sample value at that value until the next sampling time is called a sampler-clamper. Its effect is shown in Fig. **9-4.** This type of behavior can be exactly represented by the combination of an impulse modulator and a filter whose transfer function $G_e(s)$ is

$$G_e(s) = \frac{1 - e^{-st}}{s} \tag{9-1}$$

In the mathematical model, the output of the impulse modulator will be a train of impulses, the magnitude of each being the value of the actuating signal at the corresponding sampling time. Although such a signal does not exist in the physical system, it is useful to isolate the action of the impulse modulator and combine the equivalent linear filter with the output member for the purpose of analysis. The impulse modulator is thought of as a synchronous switch, controlled by a carrier, which periodically closes the connection between the actuating signal and the input to the equivalent linear filter (Fig. **9-5).** The signal $e(t)$ entering the switch is continuous in this picture, but the signal $e^*(t)$ leaving the switch is discrete (sampled). It is also possible for the signal $e(t)$ to be discrete as well, in which case the action of the switch has no effect if it is synchronized with the discrete intervals associated with the input. We will adopt the convention that an impulse modulator (or synchronous switch) operates on all

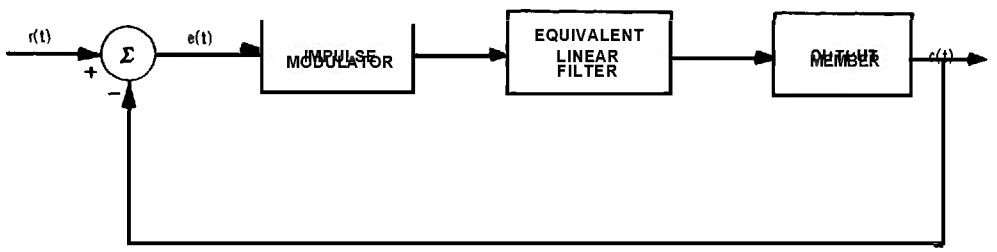


Fig. 9-2 Model of sampled-data system.

signals entering it (whether continuous, discrete, or a combination of both) to produce a discrete output. Signals that have been sampled, and are therefore discrete, will be represented by starred functions. Thus, the operation of an impulse modulator is as represented in Fig. 9-6.

The carrier signal $\Delta(t)$ in a sampled-data system is represented by

$$\Delta(t) = \sum_{n=-\infty}^{+\infty} \delta_0(t - nT) \tag{9-2}$$

The Laplace transform of this function is

$$\Delta(s) = \sum_{n=0}^{+\infty} e^{-nTs} \tag{9-3}$$

or

$$\Delta(s) = \frac{1}{1 - e^{-sT}} \tag{9-4}$$

If a signal $r(t)$ is sampled by an impulse modulator, the sampled output $r^*(t)$ is

$$r^*(t) = \Delta(t) r(t) \tag{9-5}$$

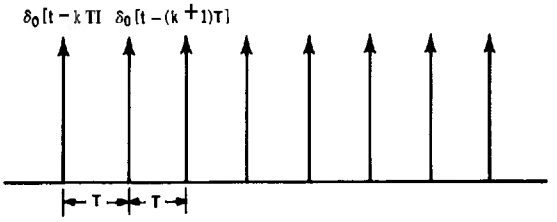


Fig. 9-3 Train of unit impulses which represents the carrier $\Delta(t)$.

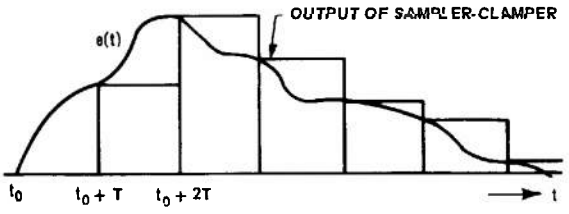


Fig. 9-4 Action of sampled-clamper.

or

$$r^*(t) = \sum_{n=-\infty}^{+\infty} r(nT)\, \delta_0(t - nT) \tag{9-6}$$

The Laplace transform $R^*(s)$ of a sampled signal is

$$R^*(s) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} R(s + jn\Omega) \tag{9-7}$$

where

$$\Omega = \frac{2\pi}{T} \tag{9-8}$$

and

$$R(s) = \mathcal{L}[r(t)] \tag{9-9}$$

Another form of the transform of a sampled signal is

$$R^*(s) = \sum_{n=0}^{\infty} r(nT)\, e^{-nTs} \tag{9-10}$$

Note that a starred transform like $R^*(s)$ represents the transform of a starred (sampled) time function. Also from Eq. (9-7), starred transforms are periodic functions of frequency, the period being $j\Omega$. That is,

$$R^*(s) = R^*(s + jn\Omega) \tag{9-11}$$

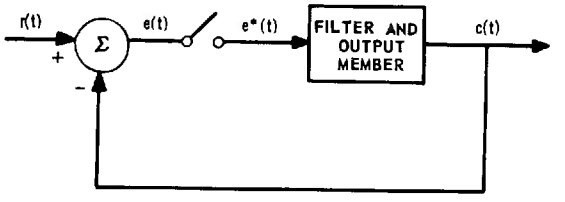


Fig. 9-5 Simplified picture of sampled-data system.

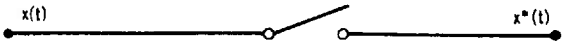


Fig. 9-6 Operation of sampling switch.

9-2 THE z TRANSFORM AND THE w TRANSFORM

9-2.1 THE z TRANSFORM

Whenever a time function is transformable, it can be shown that the transform of the function when sampled is a rational function of e^{-Ts}; i.e.,

$$R^*(s) = F(e^{sT}) \tag{9-12}$$

where $R^*(s)$ is the transform of the sampled function $r^*(t)$. If we let

$$z = e^{sT} \tag{9-13}$$

then transforms of sampled-time functions are functions of a new complex variable X. The z transform of a time function $r(t)$ is then defined as the Laplace transform $R^*(s)$ of $r^*(t)$, where $r^*(t)$ is the function produced by impulse modulating $r(t)$, and the z transform is obtained by replacing e^{sT} by z in $R^*(s)$. It is conventional to let $R^*(x)$ represent the z transform of $r(t)$ although, rigorously, one should use $R^*(-1/T \log_e z)$.

If the Laplace transform $R(s)$ of a function $r(t)$ is known, the z transform $R^*(X)$ can be found by expanding $R(s)$ in a partial-fraction expansion and using the formulae given below. If

$$R(s) = \sum_{k=1}^{m} \frac{K_k}{s + a_k} \tag{9-14}$$

then

$$R^*(z) = \sum_{k=1}^{m} \frac{K_k}{1 - ze^{-a_k T}} \tag{9-15}$$

A short table of z transforms and their equivalent Laplace transforms is given in Table 9-1. For more extensive tables see references 5, 25, and 26. Unfortunately, several authors have adopted the relation $x = e^{+sT}$. This notation arose from the mathematics of difference equations, but it is awkward and physically deceiving in the present connection, since e^{+sT} corresponds to ideal prediction. Therefore, when using the literature, care must be taken to verify which particular notation is being used. In reference 5, for example, all the expressions for X transforms should have X replaced by z^{-1} to make them correspond to the notation adopted in this chapter.

The introduction of the X transform enables one to treat sampled-data systems by all the techniques available for continuous-data systems since it is evident that the process of sampling a time function can be represented by a

TABLE 9-1
LAPLACE AND z TRANSFORM PAIRS

	Laplace Transform : $F(s)$	z Transform : $F^*(z)$
1.	1	1
2.	e^{-nTs}	z^n
3.	$\frac{1}{s}$	$\frac{1}{1-z}$
4.	$\frac{1}{s^2}$	$\frac{Tx}{(1-z)^2}$
5.	$\frac{1}{s+a}$	$\frac{1}{1-z\,e^{-aT}}$
6.	$\frac{a}{s(s+a)}$	$\frac{z(1-e^{-aT})}{(1-z)(1-ze^{-aT})}$
7.	$\frac{a}{s^2+a^2}$	$\frac{z \sin aT}{1-2z\cos uT+z^2}$
8.	$F(s+a)$	$F^*(e^{-aT}X)$
9.	$e^{-sT}F(s)$	$z\,F^*(X)$
10.	$e^{as}F(s)$	$z^{-(a/T)}F^*(x)$
11.	$\frac{1}{s+\frac{1}{T}\log_e a}$	$\frac{1}{1+ax}$
12.	$\frac{s}{s^2+a^2}$	$\frac{1-z\cos aT}{1-2z\cos aT+z^2}$
13.	$\frac{1}{(s+a)^2}$	$\frac{Te^{-aT}z}{(1-z\,e^{-aT})^2}$

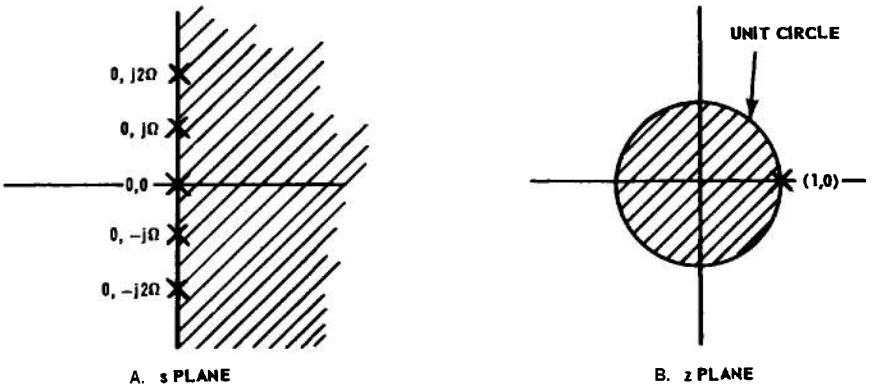


Fig. 9-7 Relations between s plane and z plane.

variable transformation from the s domain to the z domain. The z transform is a conformal transformation that maps the right half of the s plane into the interior of the unit circle of the z plane and the left half of the s plane to the exterior of the unit circle of the z plane. The imaginary axis of the s plane is mapped into the unit circle of the z plane, the s-plane origin $(0+j0)$ mapping into the point $(+1+j0)$ in the z plane. These relations are shown in Fig. 9-7. Because a z transform is periodic (period $= j\Omega$), the points $(0 + jk\Omega)$ $(k = 1, 2, 3, \ldots)$ in the s plane also map into the point $(+1+j0)$ in the z plane as shown in Fig. 9-7. Similar relationships are easily visualized and are treated later in this chapter. There is one point which must be emphasized, however. Since the z transform of a time function represents the Laplace transform of the corresponding sampled-time function, information about behavior between sampling instants is lost and cannot be recovered from inspection of z transforms. However, Barker[24] has developed a method for determining behavior between sampling instants. This method is described in Par. 9-5. It should also be noted that the z transform is related to the Mellin transform which is used to develop the theory of transforms and to study problems in elasticity.[28]

9-2.2 THE *w TRANSFORM*

Because of the difficulty involved in relating some of the properties of sampled-data systems to the frequency-domain concepts that are most convenient to apply in the study of continuous systems, Johnson et al[14] have suggested a very useful transformation of variables that aids greatly in design. If

$$z = \frac{1 - w}{1 + w} \tag{9-16}$$

is a bilateral transformation from the z plane to the w plane, then

$$R^*(w) = R^*(z) \Big|_{z=\frac{1-w}{1+w}} \tag{9-17}$$

is defined as the *w transform* of $r(t)$. The advantage of introducing the w transform becomes evident when an attempt is made to evaluate $R^*(s)$ for $s = j\omega$. Such an evaluation requires an infinite "vector" sum, theoretically [Eq. (9-7)], or else evaluation of R^* through the use of $e^{-j\omega T} = \cos \omega T + j \sin \omega T$ and so it is fairly difficult to obtain in practice. The use of the w transform, on the other hand, simplifies the determination of the frequency response of sampled-data systems. The w transform maps the unit circle in the z plane

into the imaginary axis in the w plane and restores some of the analytical advantages that were lost through the sampling process. If

$$w = u + jv \tag{9-18}$$

then

$$v = \tan\left(\frac{\omega T}{2}\right) \tag{9-19}$$

gives the relation between the real frequency ω and the pseudo- frequency v corresponding to the imaginary part of w. The real frequency ω, from Eq. (9-19), is

$$\omega = \frac{2}{T}\tan^{-1} v \tag{9-20}$$

The primary advantage of the w transform is that the transforms of sampled signals can be represented by a rational function of a frequency variable w that is simply related to the frequency ω. In the following sections this property is brought out clearly.

9-3 OPERATIONAL METHODS

9-3.1 GENERAL

The operational definition of impulse modulation given in Par. 9-1 simplifies the study of sampled-data systems.

9-3.2 BASIC RELATIONS OF SAMPLED FUNCTIONS

The following basic relations are easy to verify. From Fig. 9-8, it is evident that:

$$C(s) = G(s)R^*(s) \tag{9-21}$$

$$C^*(s) = G^*(s)R^*(s) \tag{9-22}$$

$$[R^*(s)]^* = R^*(s) \tag{9-23}$$

From Fig. 9-9, it is evident that:

$$C(s) = G(s)R(s) \tag{9-24}$$

$$C^*(s) = [G(s)R(s)]^* \tag{9-25}$$

$$C^*(s) \neq G^*(s)R^*(s)!! \tag{9-26}$$

From Fig. 9-10, it is evident that:

$$\frac{E^*(s)}{R^*(s)} = \frac{1}{1+G^*(s)} \tag{9-27}$$

$$\frac{C(s)}{R^*(s)} = \frac{G(s)}{1+G^*(s)} \tag{9-28}$$

$$\frac{C^*(s)}{R^*(s)} = \frac{G^*(s)}{1+G^*(s)} \tag{9-29}$$

The foregoing relations indicate that all the techniques of block-diagram algebra can be used to manipulate sampled-data-system configurations except for the added restrictions

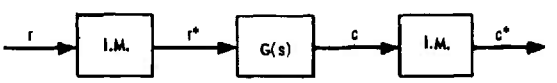


Fig. 9-8 Sampling a smoothed sampled signal.

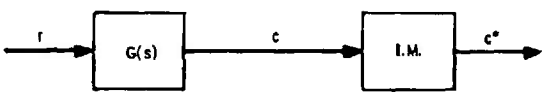


Fig. 9-9 Sampling a filtered continuous signal.

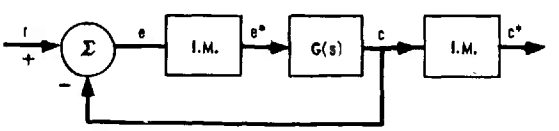


Fig. 9-10 A sampled-data feedback system.

that an impulse modulator (a) "stars" all signals entering it, and (b) its position in a diagram cannot be interchanged with a continuous transfer function. Equation **(9-26)** is included to emphasize the fact that the starred product of two Laplace transforms is ***not*** equal to the product of the corresponding starred transforms.

The equations relating to Fig. **9-10** [Eqs. **(9-27)** through **(9-29)**] introduce some of the properties of sampled-data feedback systems. In particular, the stability of a sampled-data system is related directly to the zeros of the following expression :

$$1 + G^*(s) = 0 \qquad \textbf{(9-30)}$$

If any of the roots of this equation lie in the right half of the s plane, the system is unstable. Nyquist's criterion (Par. **4-3**) can be applied directly to determine the stability of sampled-data systems, except for one modification. Since $G^*(s)$ is a periodic function, it has an infinite number of poles and zeros, but the pole-zero configuration is repeated for every multiple of $j\Omega$. Similarly, the $G^*(s)$ locus in the s plane is repeated every time s changes by $j\Omega$. Because the $G^*(s)$ plot is symmetrical about the real axis in the s plane, $G^*(s)$ need only be plotted for $0 < j\omega < j\Omega/2$ when $s = j\omega$. In practice, the s-plane contour and the corresponding $G^*(s)$ locus are as shown in Fig. **9-11** when the Nyquist criterion is applied. In terms of the z plane, Eq. **(9-30)** becomes

$$1 + G^*(z) = 0 \qquad \textbf{(9-31)}$$

The stability of the system is determined by plotting $G^*(x)$ as z traverses the unit circle. If there are any roots of Eq. **(9-31)** that lie *inside* the unit circle, the system is unstable. The difficulty encountered in plotting $G^*(s)$ or $G^*(x)$ from the required variation of s or z is removed when the w transform is introduced. $G^*(w)$ is easily handled since it is expressible as a rational function of a frequency variable.

Example. Consider a simple servomechanism with block diagram shown in Fig. **9-10.** The physical device includes a sampler-clamper (See Par. **9-1**), a servomotor having a one-second time constant, and an ideal amplifier. The transfer function of the continuous portion [including the filter as in Eq. **(9-1)**] is given by Eq. **(9-32).**

$$G(s) = K\left(\frac{1 - e^{-s}}{s}\right)\frac{1}{s(s+1)}, \quad T = 1 \text{ sec.} \qquad \textbf{(9-32)}$$

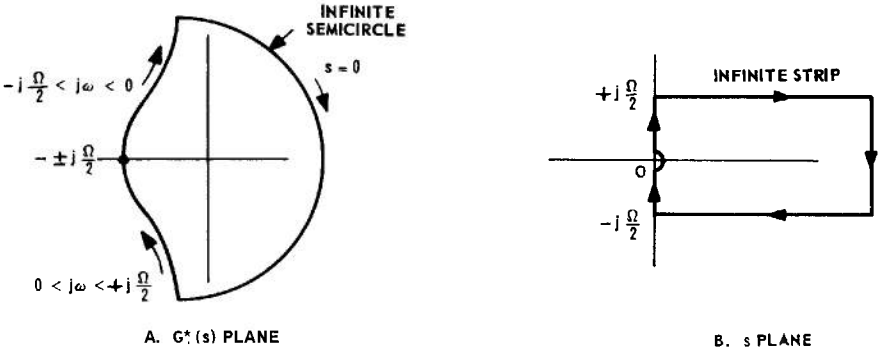


fig. 9-7 Relations between s and G(s) for application of Nyquist criterion.*

From Eq. (9-7), we obtain

$$\frac{1}{K}G^*(s) = (1-e^{-s})\sum_{n=-\infty}^{+\infty}\frac{1}{(s+jn\Omega)^2(1+s+jn\Omega)} \tag{9-33}$$

where $\Omega = 2\pi/T = 2\pi$.

This function is difficult to plot for $s = j\omega$. Taking the z transform of Eq. (9-32), we get

$$\frac{1}{K}G^*(z) = \frac{z(0.264z+0.368)}{(0.368z-1)(x-1)} \tag{9-34}$$

This function is difficult to plot for $z = e^{-j\omega}$. Letting

$$z = \frac{1-w}{1+w} \tag{9-35}$$

the w transform of Eq. (9-32) is found to be

$$\frac{1}{K}G^*(w) = \frac{(1-w)(0.632+0.104w)}{(0.632+1.368w)(2w)} \tag{9-36}$$

when w = jv, the function in Eq. (9-36) is easily handled by conventional techniques since the relation between the w plane and the $G^*(w)$ plane is the same kind of relation as that which exists between the s plane and the $G(s)$ plane. If the real frequency ω is to be considered, then there is an added difficulty in that Eq. (9-20) must be used to calibrate the frequency locus. The asymptotic and gain-phase plane techniques can be used with a change only in the relation between v and ω.

9-3.3 ADDITIONAL PROPERTIES OF SAMPLED FUNCTIONS

(a) $G^*(e^{-sT}) =$

$$\frac{1}{2\pi j}\int_{c-\infty}^{c+j\infty}G(p)\left[\frac{1}{1-e^{T(s-p)}}\right]dp \tag{9-37}$$

(b) If $G(s) = \frac{N(s)}{D(s)}$, and $D(s)$ has only simple poles,

$$G^*(e^{-sT}) = \sum_{n=1}^{k}\frac{A(s_n)}{B'(s_n)}\frac{1}{1-e^{T(s-s_n)}} \tag{9-38}$$

where s_n is the nth pole of $G(s)$.

(c) If the z transform of a time function $r(t)$ is given, the value of the time function at the sampling instants can be found from

$$r^*t = r(nT) = \frac{1}{2\pi j}\oint R^*(z)z^{1-n}\,dz \tag{9-39}$$

where the contour integration in the z plane is along a path that encloses all the singular points of $R^*(x)z^{1-n}$.

(d) Initial-Value Theorem:

$$\lim_{t\to 0} r(t) = \lim_{z\to 0}(1-z)R^*(z) \tag{9-40}$$

(e) Final-Value Theorem: If $R^*(x)$ has all its poles outside the unit circle of the z plane,

$$\lim_{t\to\infty} r(t) = \lim_{z\to 1}(1-z)R^*(z) \tag{9-41}$$

(f) For $s = \pm j\frac{k\Omega}{2}(k = 0, 1, 2, \ldots)$, $G^*(z)$ is always real $\left(\Omega = \frac{2\pi}{T}\right)$.

(g) The degree of the denominator of $G^*(x)$ in z always equals the degree of the denominator of $G(s)$ in s if $G(s)$ is rational.

(h) The *poles* of $G^*(x)$ in the strip $-j\frac{\Omega}{2} < Im(s) < +j\frac{\Omega}{2}$ in the splane are the *poles* of $G(s)$ in the s plane.

(i) Changing the values of the poles of $G(s)$ changes the coefficients $A(s_n)/B'(s_n)$ as well as the terms $1/(1-ze^{+Ts_n})$ in the partial-fraction expansion of $G^*(z)$ [Eq. (9-38)].

(j) Insertion of zeros in $G(s)$ changes only the coefficients $A(s_n)/B'(s_n)$.

(k) The number of poles of $G^*(x)$ at $z = 1$ is equal to the number of poles of $R(s)$ at $s = 0$.

(l) In terms of the w transform, the initial value theorem is

$$\lim_{t\to 0} r(t) = \lim_{w\to 1}\left(\frac{2w}{1+w}\right)R^*(w) \tag{9-42}$$

and the final-value theorem is

$$\lim_{t\to\infty} r(t) = \lim_{w\to 0}\left(\frac{2w}{1+w}\right)R^*(w) \tag{9-43}$$

9-4 DESIGN TECHNIQUES

The problem of designing a sampled-data system is complicated by the fact that the system can contain both discrete and continuous elements. In addition, direct application of x-transform theory merely gives the response at the sampling instants, but the behavior during the sampling instants cannot be determined by simple methods.

The insertion of a sampling device in an otherwise continuous system to produce a sampled-data system introduces the following limitations:

(a) A greater tendency towards instability results.

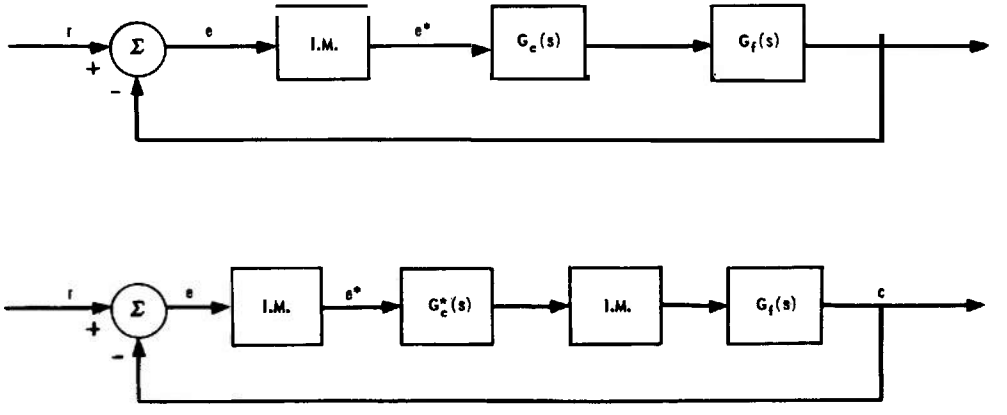
(b) Ripple components arise in the output at the sampling frequency and its harmonics.

(c) The usable bandwidth of the system is reduced to a fraction of the sampling frequency Ω, the theoretical upper limit being $\Omega/2$.

To determine the gain necessary to stabilize the system for a specified M_p in the closed-loop frequency response, the introduction of the w transform greatly facilitates plotting the frequency locus as indicated in Par. 9-3. Conventional continuous-system techniques can be used.

The root-locus procedure can be used in a conventional manner in the *z plane* to investigate the closed-loop pole-zero configuration. This procedure differs from that used for s-plane loci of continuous systems in that: (a) instability implies closed-loop poles inside the unit circle of the z plane (as contrasted to right-half-plane poles in continuous-system design), and (b) the dominant pole (or pole pair) is the pole nearest the point (1,0) in the z plane (as contrasted to poles nearest the origin in the s plane). Otherwise, conventional procedures can be used to investigate stability, relative stability, and the effect of compensation.

Compensation of sampled-data systems with continuous networks (conventional lead and lag networks) is a difficult design problem and is best treated by trial-and-error analysis. In many important applications discrete networks can be used for compensation; for example, the use of digital computers in fire-control systems provides the designer with an opportunity to use digital (discrete) filters in the compensation of the control system. Figure 9-12 shows the difference between continuous and discrete com-



8. DISCRETE COMPENSATION

Fig. 9-12 Comparison between discrete and continuous compensation.

pensation. In Fig. 9-12A, $G_c(s)$ is the transfer function of a continuous network which is used to improve the closed-loop behavior; $G_f(s)$ is the transfer function of the fixed elements. In Fig. 9-12B, $G_c^*(s)$ is the transfer function of a digital network (digital program) used to improve system performance. For the case of continuous compensation, we have

$$\frac{E^*(s)}{R^*(s)} = \frac{1}{1 + [G_c(s)G_f(s)]^*} \tag{9-44}$$

$$\frac{C(s)}{R^*(s)} = \frac{G_c(s)G_f(s)}{1 + [G_c(s)G_f(s)]^*} \tag{9-45}$$

For digital compensation, we have

$$\frac{E^*(s)}{R^*(s)} = \frac{1}{1 + G_f^*(s)G_c^*(s)} \tag{9-46}$$

$$\frac{C(s)}{R^*(s)} = \frac{G_f(s)G_c^*(s)}{1 + G_f^*(s)G_c^*(s)} \tag{9-47}$$

Compensation with a digital network is a simpler analytical problem than continuous compensation because the effect of the compensation can be varied without altering the fixed-element contribution to the open-loop response $G_f^*(s)G_c^*(s)$. For continuous compensation, the open-loop response is $[G_c(s)G_f(s)]^*$, and altering the compensation will alter the contribution of the fixed elements to the open-loop response. Conventional continuous system techniques can be used for synthesizing digital programs to compensate sampled-data systems ;but if continuous compensation is desired, a trial-and-error analytical procedure is necessary.

9-5 PERFORMANCE EVALUATION

The determination of the time response of a sampled-data system can be carried out in closed form by the use of Eqs. (9-38) and (9-39). A simple numerical procedure for determining the values of the output at the sampling instants can be obtained by expanding the z transform of the output in a power series in z. Since z corresponds to a time delay of one sampling instant, the coefficients of the power-series expansion of a z transform are the values of the corresponding time function at the sampling instants, as can be seen from an examination of Eq. (9-10). The expansion is easily performed by dividing the numerator by the denominator since the z transform is a rational function.

Example.

Assume that the z transform of the output is given by:

$$C^*(z) = \frac{0.186(z^2 + 1.392)}{(1-x)[0.554z^2 - 1.108z + 1]}$$

Dividing the numerator by the denominator, the power series expansion of $C^*(x)$ is found to be

$$C^*(z) = (.26)z + (.76)z^2 + (1.17)z^3 + (1.36)z^4 + (1.33)z^5 + (1.20)z^6 + (1.07)z^7 + (.99)z^8 + (.96)z^9 + \ldots$$

The coefficients of this expansion are the sampled values of $c(t)$, and the instant of occurrence is determined from the power of z in the appropriate term. The function is plotted in Fig. 9-13.

An alternate method for finding the time response of a system is based on a difference-equation representation. Assume that

$$\frac{C^*(s)}{R^*(s)} = \frac{a_0 + a_1e^{-sT} + a_2e^{-2sT} + \ldots}{b_0 + b_1e^{-sT} + b_2e^{-2sT} + \ldots} \tag{9-48}$$

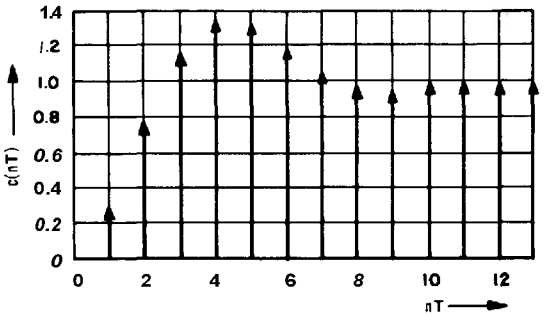


Fig. 9-13 Sampled-time function.

Cross-multiplying, inverse transforming, and solving for $c(nT)$, we obtain

$$c(nT) = \frac{1}{b_0}[a_0 r(nT) + a_1 r[(n-1)T] + \ldots - b_1 c[(n-1)T] - b_2 c[(n-2)T] - \ldots \quad (9\text{-}49)$$

This is a general recurrence formula which enables one to calculate the present value $c(nT)$ of the output in terms of a weighted sum of the present and past values of the input $r(t)$ and the past values of the output. The calculation is best carried out in tabular form.

Example. Assume that the closed-loop transfer function of a sampled-data servomechanism is

$$\frac{C^*(z)}{R^*(z)} = \frac{1.5z}{1+0.5z}$$

If $r(t)$ is a unit step, the tabular evaluation of $c(nT)$ can be carried out as follows. Cross-multiplying, we get

$$C^* + 0.5e^{-aT}C^* = 1.5e^{-aT}R^*$$

Inverse transforming, we obtain

$$c(nT) + .05\,c[(n-1)T] = 1.5r[(n-1)T]$$

or

$$c(nT) = 1.5r[(n-1)T] - 0.5c[(n-1)T]$$

when $r(nT)$ is a unit step, and where the system has been at rest so that $c(-T)$ is zero, the calculation of $c(nT)$ can be carried out in tabular form as follows:

n	$1.5r$ $[(n-1)T]$	$-0.5c$ $[(n-1)T]$	$c(nT)$
0	0	0	
1	1.5	0	0
2	1.5	−0.75	1.5
3	1.5	−0.375	0.75
4	1.5	−0.562	1.125
5	1.5	−0.469	0.938
6	1.5	−0.516	1.031
7	1.5	−0.492	0.984
8	1.5	−0.504	1.008
9	1.5	−0.498	0.996
			1.002

The values in the last column are plotted in Fig. 9-14. In general, it is necessary to know the value of $c(nT)$ for values of n corresponding to time prior to the beginning of the transient. If the system is of kth order, it is necessary to know $c(nT)$ for k prior to the samples. This corresponds to the need for knowing the initial conditions in any transient problem.

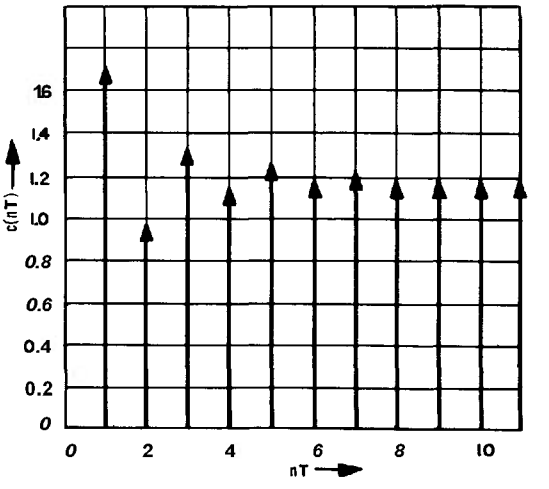


Fig. 9-14 Step response of sampled-data system.

Since the z transform does not give the value of the output between sampling instants, Lago and **Truxal**[10] have suggested a method which enables one to determine the output at submultiples of the sampling period T. A fictitious impulse modulator is placed immediately after the actual one (Fig. **9-15).** The fictitious sampler has a sampling period which is an integral submultiple of T; i.e.,

$$T' = \frac{T}{n} \tag{9-50}$$

As a result, we can write

$$C'^*(z) = \frac{R^*(z^n)}{1 + G^*(z^n)}\, G'^*(z) \tag{9-51}$$

where $G'^*(x)$ and $C'^*(z)$ are z transforms **of** $g(t)$ and $c(t)$ with respect to the period T', and $R^*(z^n)$ and $G^*(z^n)$ are z transforms of $r(t)$ and $g(t)$ with respect to the period T but with z replaced by z^n.

An extension of the method above is given by **Barker.**[24] In this extended method, the output at *any* time between sampling instants can be found. If we refer to Fig. **9-16,** it can be seen that the artificial delay produces a signal **c'** (t) which one can sample in order to observe the values which will occur between the values of $c^*(t)$. We find that

$$C'^*(z,m) = \frac{G^*(z,m)}{1 + G^*(x)}\, R^*(z) \tag{9-52}$$

where $G^*(z,m)$ is a modified z transform. $G^*(z,m)$ is evaluated by assuming that $g(t)$ is sampled at $t = (n + m - 1)T$ instead of at $t = nT$. A brief table of modified z transforms is listed in Table **9-2.** A more extensive table is given by **Barker.**[24] The use **of** Eq. (9-52) enables one to *scan* the output by varying m between zero and unity. Thus, the variation **of** the output between sampling instants is observed, and a study of the ripple can be made.

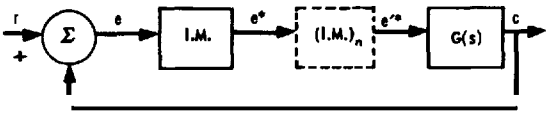


Fig. 9-75 Determination of c(t) between sampling instants by sampling at n Ω rad/sec.

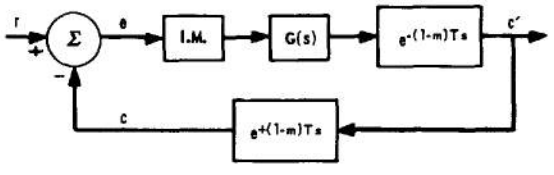


Fig. 9-76 Determination of c(t) between sampling instants through the use of an artificial delay.

The ***error*** *coefficients* of a sampled-data system may be obtained from the expression

$$e_n = \frac{1}{n!}\left\{\frac{d^n}{ds^n}\left[\frac{E^*(s)}{R^*(s)}\right]\right\}_{s=0} \tag{9-53}$$

Since $E^*(s)/R^*(s)$ is a rational function in z, where $z = e^{sT}$, we can expand it in a Taylor series (in x) about the point $x = 1$ $(s = 0)$. Then, by using the infinite series expansion of e^{sT} and rearranging terms, we can easily obtain the Taylor series expansion of $E^*(s)/R^*(s)$ in terms of s at $s = 0$:

$$\frac{E^*(z)}{R^*(z)} = a_0 + a_1(z-1) + a_2(z-1)^2 + a_3(z-1)^3 + \ldots \tag{9-54}$$

or

$$\frac{E^*(s)}{R^*(s)} = a_0 - T(a_1)s + T^2(a_2 + \frac{a_1}{2!})s^2 - T^3(a_3 + a_2 + \frac{a_1}{3!})s^3 + \cdots \quad (9\text{-}55)$$

The techniques described in this section can be used to obtain the transient response, the ripple, and the error coefficients of a sampled-data system. Using trial-and-error procedures and the conventional continuous-system design techniques, the evaluation of a given system in terms of a set of performance specifications is a straightforward matter.

TABLE 9-2 MODIFIED z TRANSFORMS

$F(s)$	$F^*(z,m)$
1	0
$\frac{1}{s}$	$\frac{z}{1-z}$
$\frac{1}{s^2}$	$\frac{Tz^2 + mz}{(1-z)^2}$
$\frac{1}{s+a}$	$\frac{ze^{-amT}}{1-ze^{-aT}}$

BIBLIOGRAPHY

1 W. K. Linvill, "Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation", *Trans. AIEE,* Vol. 70, Part 11, pp. #1779-1788, 1951.

2 K. S. Miller and R. J. Schwarz, "Analysis of a Sampling Servomechanism", J. *Appl. Phys.,* Vol. 21, pp. #290-294, 1950.

3 J. R. Ragazzini and L. A. Zadeh, "The Analysis of Sampled-Data Systems", *Trans. AIEE,* Vol. 71, Part II, pp. #225-234, 1952.

4 R. G. Brown and G. J. Murphy, "An Approximate Transfer Function for the Analysis and Design of Pulsed Servos", *Trans. AIEE,* Vol. 71, Part II, pp. #435-440, 1952.

5 E. I. Jury, "Analysis and Synthesis of Sampled-Data Control Systems", *Trans. AIEE,* Vol. 73, Part I, pp. #332-346, 1954.

6 W. K. Linvill and J. M. Salzer, "Analysis of Control Systems Involving Digital Computers", Proc. *IRE,* Vol. 41, pp. #901-906, July, 1953.

7 A. R. Bergen and J. R. Ragazzini, "Sampled-Data Processing Techniques for Feedback Control Systems", *Trans. AIEE,* Vol. 73, Part 11, pp. #236-247, 1954.

8 C. K. Chow, "Contactor Servomechanisms Employing Sampled Data", *Trans. AZEE,* Vol. 73, Part 11, pp. #51-64, 1954.

9 G. W. Johnson and D. P. Lindorff, "Transient Analysis of Sampled-Data Control Systems", *Trans. AIEE,* Vol. 73, Part 11, pp. #147-153, 1954.

10 G. V. Lago and J. G. Truxal, "The Design of Sampled-Data Feedback Systems", *Trans. AIEE,* Vol. 73, Part 11, pp. #247-253, 1954.

11 G. V. Lago, "Additions to z-Transformation Theory for Sampled-Data Systems", *Trans. AIEE,* Vol. **73,** Part 11, pp. #403-408, **1954.**

12 E. I. Jury, "The Effect of Pole and Zero Locations on the Transient Response of Sampled-Data Systems", *Trans. AIEE,* Vol. **74,** Part 11, pp. #41-48, **1955.**

13 J. Sklansky and J. R. Ragazzini, "Analysis of Errors in Sampled-Data Systems", *Trans. AIEE,* Vol. **74,** Part 11, pp. #65-71, **1955.**

14 G. W. Johnson, D. P. Lindorff, and C. G. A. Nordling, "Extension of Continuous-Data System Design Techniques to Sampled-Data Control Systems", *Trans. AIEE,* Vol. **74,** Part 11, pp. #252-263, **1955.**

15 E. I. Jury, "Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems", *Trans. AIEE,* Vol. **74,** Part II, pp. #427-435, **1955.**

16 K. K. Maitra and P. E. Sarachik, "Digital Compensation of Continous-Data Feedback Control Systems," *Trans. AIEE,* Vol. **75,** Part 11, pp. #107-116, **1956.**

17 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory* of *Servomechanisms,* MIT Radiation Laboratory Series, Vol. **25,** pp. #231-261, McGraw-Hill Book Co., Inc., New York, N. Y., **1947.**

18 H. S. Tsien, *Engineering Cybernetics,* pp. #83-93, McGraw-Hill Book Co., Inc., New York, N.Y., **1954.**

19 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #500-558, McGraw-Hill Book Co., Inc., New York, N.Y., **1955.**

20 **B.** M. Brown, "Application of Finite Difference Operators to Linear Systems", in *Automatic and Manual Control,* edited by A. Tustin, pp. #409-418, Butterworths Scientific Publications, London, England, **1953.**

21 C. Holt Smith, D. F. Lawden, and A. E. Bailey, "Characteristics of Sampling Servo Systems", in *Automatic and Manual Control,* edited by A. Tustin, pp. #377-404, Butterworths Scientific Publications, London, England, **1952.**

22 R. C. Oldenbourg, "Deviation Dependent Step-by-step Control as Means to Achieve Optimum Control for Plants with Large Distance-Velocity Lag", in *Automatic and Manual Control,* Proceedings of Cranfield Conference, **1951,** edited by A. Tustin, pp. #435-447, Butterworths Scientific Publications, London, England, **1952.**

23 H. Sartorius, "Deviation Dependent Step-by-Step Control Systems and Their Stability", in *Automatic and Manual Control,* Proceedings of Cranfield Conference, 1951, edited by A. Tustin, pp. #421-434, Butterworths Scientific Publications, London, England, **1952.**

24 R. H. Barker, "The Pulse Transfer Function and Its Application to Sampling Servosystems", ***Proc.*** *IEE* (London), Vol. **99,** Part IV, **1952.**

25 W. M. Stone, "A List of Generalized Laplace Transforms", *Journal* of *Science,* Iowa State College, Ames, Iowa, Vol. **22,** pp. #215-225, April, **1948.**

26 M. F. Gardner and J. L. Barnes, *Transients in Linear Systems,* Vol. I, pp. #354-356, John Wiley & Sons, Inc., New York, N.Y., **1942.**

27 D. F. Lawden, **"A** General Theory of Sampling Servosystems", ***Proc.*** *IEE* (London), Vol. **98,** Part **IV,** pp. #31-36, October, **1951.**

28 I. N. Sneddon, *Fourier Transforms,* McGraw-Hill Book Co., Inc., New York, N.Y., **1951.**

CHAPTER 10

NONLINEAR SYSTEMS*

10-1 INTRODUCTION

All of the techniques of system analysis discussed in previous chapters are restricted in their application to linear systems. This restriction imposes two limitations on design. First, components must be of high quality if they are to operate in a linear manner when amplitudes and frequencies of signals vary widely. Second, the linearity restriction limits the realizable system characteristics, the types of systems, and the tasks that can be accomplished.

Nonlinearities are generally of two types: incidental and intentional. Incidental nonlinearities are secondary effects which limit performance in otherwise linear systems. Examples of phenomena that introduce incidental nonlinearities include backlash, saturation, dead zone, hysteresis, and coulomb friction. On the other hand, intentional nonlinearities are those introduced purposely to improve the characteristics of systems or to alter them in specified ways. The contactor (on-off or relay) servo is the most extreme example of such an intentionally nonlinear system.

10-2 DESCRIBING FUNCTION PROCEDURES[7,8,12,13,15,18,19,21,22,31,35,36,42,50,51,52]

One problem to be analyzed in an investigation of nonlinear system behavior is that concerned with the question of stability. A method of studying this problem utilizes the describing-function procedure. The application of the *describing-function* procedure enables the designer to predict whether or not a closed-loop system containing a nonlinear element will be stable. A system is said to be stable if, after a sudden input or disturbance, it eventually comes to rest. In some systems the existence of a stable oscillation is acceptable provided the amplitude of the oscillation is small. A typical case is a relay control system where a small amplitude oscillation may be acceptable if the cost of eliminating the oscillation is too high. A system is said to be unstable if a finite input or disturbance to the system results in an output oscillation that tends to grow without bound.

The describing-function method is based on three assumptions:

(a) There is only one nonlinear element in the system. If there are more than one, that part of the system including all nonlinearities is treated as a single nonlinear component.

(b) The characteristics of the nonlinear element are independent of time. They depend only on the present value and past history of the input to the element.

(c) If the input to the nonlinear element is sinusoidal, only the fundamental sinusoidal component of the output of the element contributes to the input of this element.

*By L. A. Gould

The last assumption is the heart of the method. It is applicable when the amplitude of the harmonics generated by the nonlinearity decreases and when the elements that follow the nonlinearity have low-pass characteristics.

Referring to Fig. **10-1,** the method of analysis is as follows. The input $x(t)$ to the nonlinear element N is assumed to be sinusoidal with amplitude X and frequency ω. The output $y(t)$ will be periodic (but nonsinusoidal) with the same frequency ω. A Fourier analysis of the output waveform is made and all frequencies except the fundamental are ignored. The amplitude Y, of the fundamental will, in general, be a function of X and ω, as will be the phase angle of the fundamental relative to the input. The describing function of the nonlinear element is defined as a complex number whose magnitude is Y_1/X, the ratio of the amplitude of the fundamental component of the output to the amplitude of the input, and whose angle is the phase angle of the fundamental component of the output relative to the phase angle of the input. The describing function is usually denoted by $N(X,\omega)$. Symbolically, if

$$x(t) = X \sin \omega t \tag{10-1}$$

then

$$y(t) = Y_1 \sin(\omega t + \phi_1) + Y_2 2 \sin(\omega t + \phi_2) + Y_3 3 \sin(\omega t + \phi_3 + \ldots \tag{10-2}$$

and

$$\left| N(X,\omega) \right| = \frac{Y_1}{X} \tag{10-3}$$

$$\angle N(X,\omega) = \phi_1 \tag{10-4}$$

The describing functions of several important nonlinearities will now be presented. Figure **10-2** shows the input-output characteristic of a ***contactor*** with inactive zone A (dead zone) and hysteresis ***h.*** The amplitude and phase curves of the describing function N of

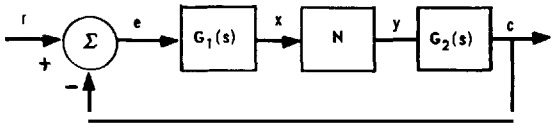


fig. 10-1 Nonlinear feedback control system.

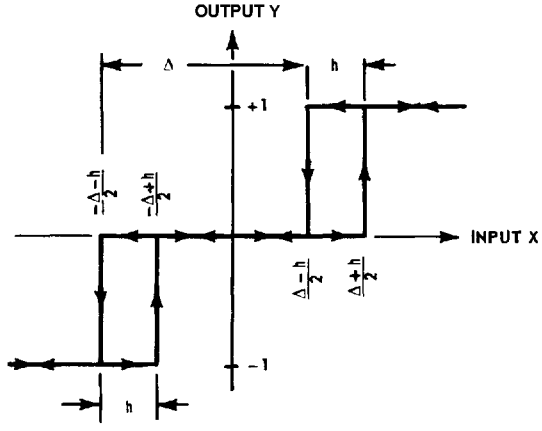


Fig. 10-2 Dimensionless representation of contactor characteristics (case involving both inactive zone and hysteresis).

Adapted by permission from *Transactions of the AIEE.* Volume 69, Part I, 1950, from article entitled 'A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms', by R. J. Kochenburger.

this contactor as functions of the input amplitude X appear in Fig. 10-3. Figure 10-4 shows the input-output characteristics of a *nonlinear element containing both dead zone D and saturation S.* No phase shift is associated with this element since, in general, phase shift will not occur for a single-valued nonlinearity. The describing function for no saturation ($S \rightarrow \infty$) is presented in Fig. 10-5. The describing function for no dead zone ($D = 0$) is presented in Fig. 10-6. Describing functions for various combinations of dead zone and saturation appear in Fig. 10-7. Figure 10-8 shows the input-output characteristic of a nonlinear element characterized by *hysteresis* (backlash or free play). The describing function for this nonlinearity is presented in Fig. 10-9. Other describing functions for more complex nonlinearities can be derived for the particular case being considered. Additional describing functions are given in the literature.[18,19,22]

The procedure for using the describing function to predict the nature of the stability of a nonlinear system follows. Referring to

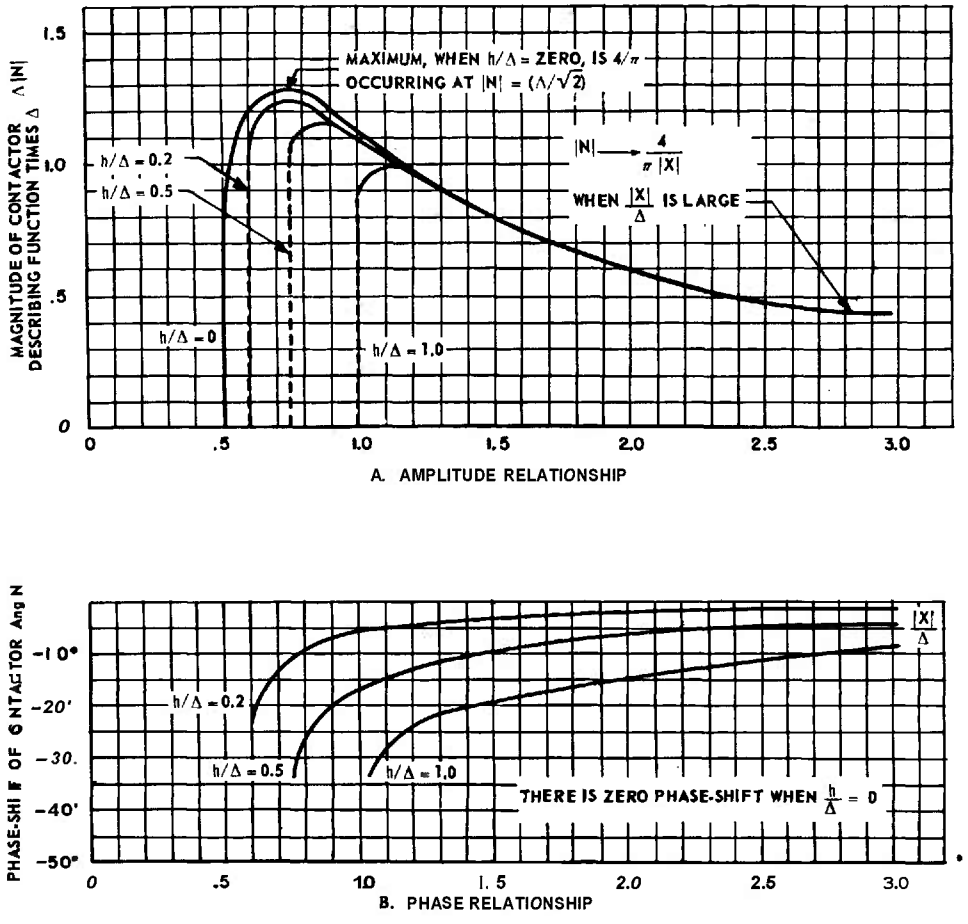


Fig. 10-3 Plot of the describing function N (simple contactor with hysteresis ratio h/Δ).

Adapted by permission from *Transactions of the AIEE,* Volume 69, Part I, 1960. from article entitled 'A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms', by R. J. Kochenburger.

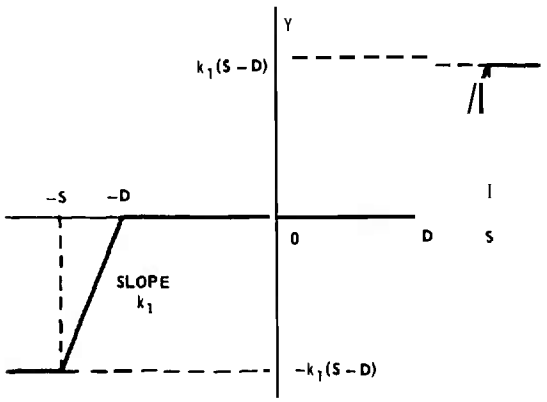


Fig. 10-4. Nonlinear characteristic with dead zone and saturation.

By permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal, Copyright, **1955,** McGraw-Hill Book Company, Inc.

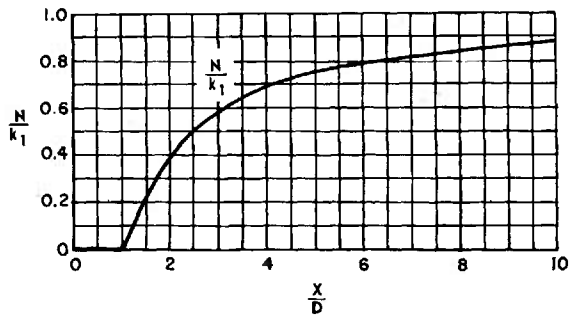


Fig. 10-5 Describing function for dead zone.

Adapted by permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal, Copyright, **1955,** McGraw-Hill Book Company, Inc.

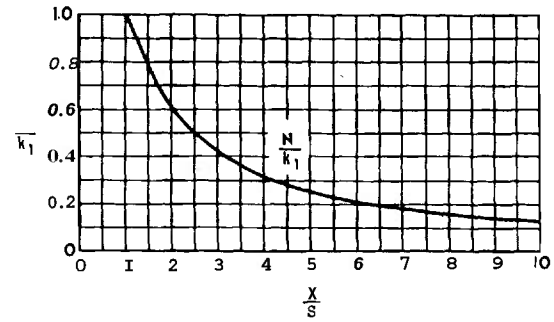


Fig. 10-6 Describing function for saturation.

Adapted by permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal, Copyright, **1955,** McGraw-Hill Book Company, Inc.

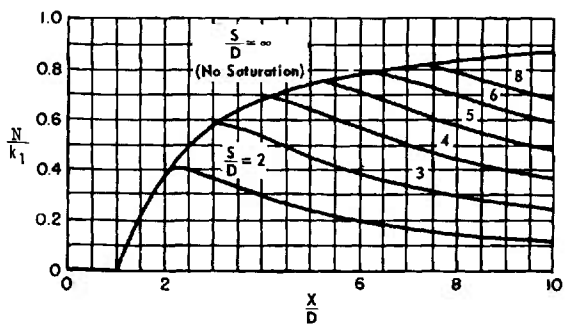


Fig. 10-7 Describing function for saturation and dead zone.

By permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal, Copyright, **1955,** McGraw-Hill Book Company, Inc.

Fig. 10-10, the linear and nonlinear portions of the system are separated into two parts; the describing function N applies to one part, and the response of the linear elements G to the other. The *gain-phase plane* is employed for plotting the negative reciprocal $(-1/N)$ of the describing function. The response $G(j\omega)$ of the linear elements is also plotted on the same plane. If the $-1/N$ locus and the $G(j\omega)$ locus do not intersect, the system is stable and does not oscillate. If the $-1/N$ locus and the $G(j\omega)$ locus do intersect (two types of intersections can occur), the system may or may not be oscillatory. The describing function for a contactor with hysteresis and dead zone is sketched in Fig. 10-11 wherein the types of intersections of the $-1/N$ locus with a $G(j\omega)$ locus are shown. The parameter along the $-1/N$ locus is the amplitude X of the assumed sinusoidal input tu the nonlinear

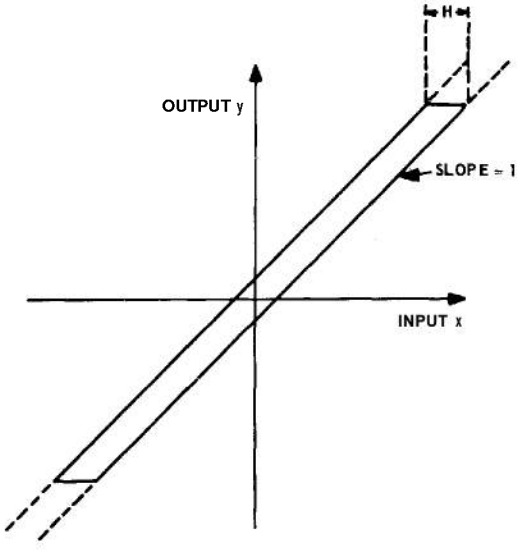


Fig. 10-8 Hysteresis nonlinearity.

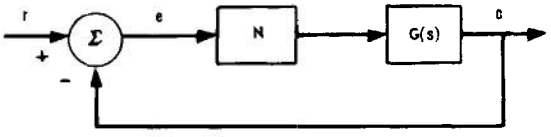


Fig. 10-10 Simplified nonlinear system.

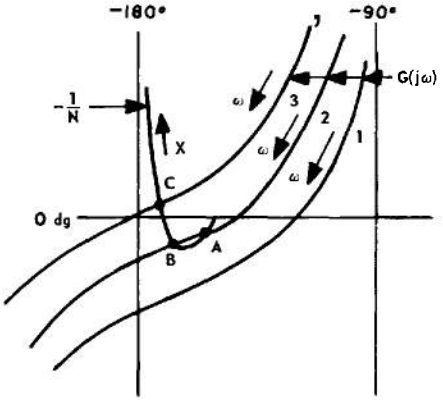


Fig. 10-11 Stability determination with describing function.

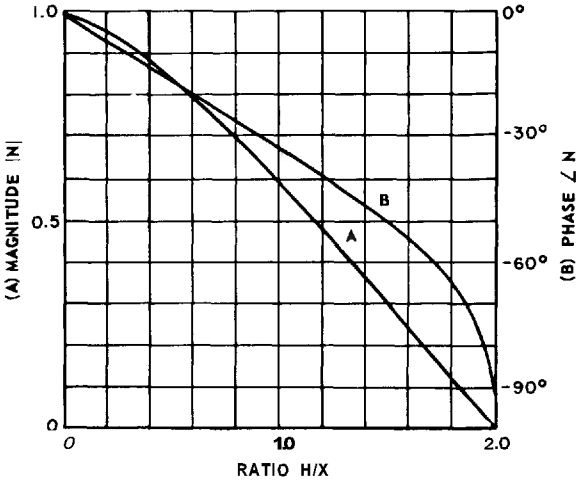


Fig. 10-9 Describing function for hysteresis-type nonlinear element.

Adapted by permission from *Transactions of the AIEE*. Volume 72, Part II, 1953, from article entitled 'Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities'. by H. D. Grief.

element. Three cases illustrating the types of intersection of the $G(j\omega)$ locus with the $-1/N$ locus are shown in Fig. 10-11.

Case 1. The $G(j\omega)$ locus does not intersect the $-1/N$ locus. The system is stable and no oscillation occurs.

Case **2.** The $G(j\omega)$ locus intersects the $-1/N$ locus at two points A and B. Point **A** is called a *divergent equilibrium* point since sustained oscillations cannot be maintained at the frequency ω_A and amplitude X_A associated with A. The existence of a divergent equilibrium can be determined by treating the $-1/N$ locus as one treats the $-1 + j0$ point in the study of the stability of linear systems. If the amplitude X associated with point A decreases slightly, the $G(j\omega)$ locus will be located in a stable position with respect to the $-1/N$ point, and oscillations will tend to die out. If the amplitude X tends to increase from X_A, the $G(j\omega)$ locus encloses the $-1/N$ point

(corresponding to instability), and the amplitude of oscillation will tend to increase. Thus, oscillation cannot be maintained at A. However, point B is a point of *convergent equilibrium,* as can be determined by letting the amplitude X both increase and decrease relative to X_B. In each case, the tendency will be for the amplitude of the oscillation to head back to point B. Thus, the convergent equilibrium point B determines an amplitude X_B (read off the $-1/N$ locus) and a frequency ω_B [read off the $G(j\omega)$ locus] at which a sustained oscillation occurs.

Case 3. The $G(j\omega)$ locus intersects the $-1/N$ locus at point C. This is a *convergent equilibrium* point, as can be determined by letting the amplitude X increase or decrease relative to the amplitude X_c associated with the intersection C. In each case, the amplitude of oscillation will tend to return to X_c.

The describing-function method thus predicts the stability of nonlinear systems, as described above. If intersections occur between the $-1/N$ locus and the $G(j\omega)$ locus, the amplitude and frequency of convergent oscillations can be predicted to an accuracy that is determined by the assumptions inherent in the method. Techniques for estimating the accuracy of the results are given in references 15 and 52.

The describing-function procedure breaks down if the two loci ($-1/N$ and G) approach each other without intersecting, are tangent, or intersect at a small angle. In these situations one cannot be certain as to whether or not oscillations exist.

In general, the accuracy of the describing-function method increases as the cutoff rate of the linear element *(following* the nonlinear element) increases. The accuracy may decrease if the linear element exhibits a sharp resonance.

The describing-function procedure is useful in predicting the closed-loop frequency response of a system containing an incidental nonlinearity when no oscillation can occur. Thereby, peculiarities in measured characteristics can be explained, and quantitative estimates of nonlinear effects can be made. By treating the $-1/N$ locus as the equivalent of the $-1 + j0$ point of conventional linear analysis, the degree of stability of a system containing a nonlinear element may be estimated. Instead of aligning the (−180°, 0 dg) point of the Nichols chart with the (−180°, 0 dg) point of the $G(j\omega)$ locus in order to determine M_p, as is done when a linear element is present, the (−180°, 0 dg) point of the Nichols chart is aligned with a point chosen on the $-1/N$ locus of the nonlinear element for a given amplitude X of the input to the element, when a nonlinear element is present. The tangency of the G locus (plotted on the same coordinates as the $-1/N$ locus) to an M contour of the Nichols chart will then be an indication of the degree of stability associated with the chosen amplitude X. Moving the (−180°, 0 dg) point of the Nichols chart along the $-1/N$ locus is equivalent to changing the amplitude X of the input to the nonlinear element. By this means, the variation of the degree of stability (as measured by M_p) can be determined as a function of the amplitude of the input to the nonlinear element. The relation between the amplitude X (input to nonlinear element) and the reference-input amplitude R can be determined for each M_p value from the following relations (see Fig. 10-1).

$$|R| = \frac{|N|\ |G_2(j\omega_p)|}{M_p}\,|X| \qquad \textbf{(10-5)}$$

where ω_p is the frequency associated with the point of M_p tangency for each value of X along the $-1/N$ locus.

These methods can be extended to determine the entire frequency response of a system by noting the intersections of other M contours with the G locus at each value of X and using the following relations to determine the input amplitude (or amplitudes) R associated with each value of X:

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| = \left|\frac{G_1(j\omega)N\,G_2(j\omega)}{1 + G_1(j\omega)N\,G_2(j\omega)}\right| \qquad (10\text{-}6)$$

$$\left|\frac{X(j\omega)}{R(j\omega)}\right| = \frac{1}{|NG_2(j\omega)|}\left|\frac{C(j\omega)}{R(j\omega)}\right| \qquad \textbf{(10-7)}$$

Example. A relay (contactor) servomechanism employs a relay with a ratio of hysteresis h to dead zone A of 0.5. The describing function of the relay is plotted in Fig. **10-12** as a function of the normalized input amplitude a, where $a = X/\Delta$. The linear element $G_1(j\omega)$ is a pure gain K (see Fig. **10-1**). The linear element $G_2(j\omega)$ is represented by the relation

$$G_2(j\omega) = \frac{1}{j\omega(0.05j\omega + 1)}$$

The block diagram of the system is shown in Fig. **10-13.** The response G_1G_2 of the linear elements is plotted in Fig. **10-12** with $K/A = 20$ as curve (B). No intersection occurs, and so the system is stable. If the gain factor K is increased by 8.5 dg (a factor of **7.1**), curve (A) is obtained. Now, an intersection occurs for a = 0.9 and $\omega = 55$ rad/sec. This intersection is a convergent equilibrium point, and the system will therefore oscillate at **55** rad/sec with an error amplitude $E = 0.00635$ since

$$|E| = \frac{a}{[K/\Delta]}$$

and

$$K/A = 142$$

For the stable case, curve (B) of Fig. **10-12** and Eq. **(10-5)** are used to determine the variation of M_p with the magnitude $|R|$ of the sinusoidal input. The results are plotted in Fig. **10-14.**

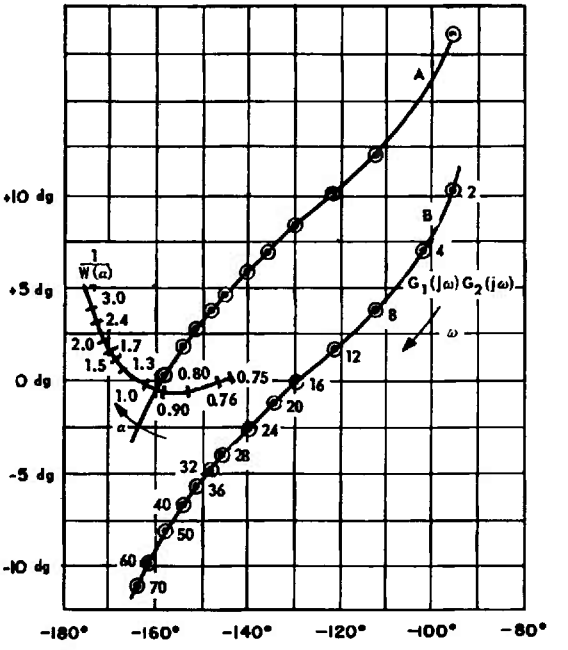


Fig. 10-12 Contactor servomechanism study.

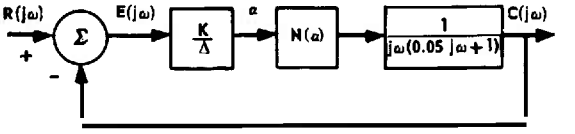


Fig. 10-13 Contactor servomechanism.

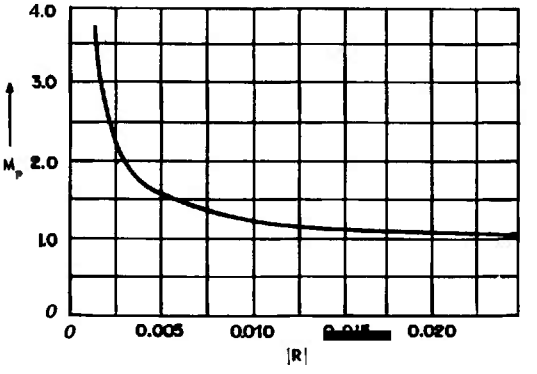


Fig. 10-14 Degree of stability variation with input amplitude for contactor servomechanism.

10-3 PHASE-PLANE PROCEDURES[2,3,4,5,6,23,24,25,26,27,30,34,39,40,41,44,46,47,48,51,52,53]

The main limitation of the describing-function procedure is that it cannot predict the response of stable nonlinear systems to inputs that are not sinusoidal. In contrast, the *phase-plane method* is an attempt to describe the response of nonlinear systems to specific transient inputs. In this method, attention is focused on the differential equations that describe the system, and the behavior of the system is studied by plotting velocity versus displacement with time as a parameter. This velocity-displacement plane is called the *phase plane*. Only second-order systems can be handled in the phase plane although attempts have been made to treat higher-order systems by a phase-space representation.[34,40,46,47,48,53]

Phase-plane analysis is concerned with the characteristics of the differential equation

$$\ddot{x} + a(x,\dot{x})\dot{x} + b(x,\dot{x})x = 0 \qquad (10\text{-}8)$$

The phase-plane portrait of the system is a plot of the velocity $\dot{x}$ as a function of the displacement x, the plot being a family of curves depending on the initial conditions $x(0)$ and $\dot{x}(0)$. Once the initial conditions have been specified, the behavior of the system is determined completely by the curve in the phase plane corresponding to the given initial conditions. Thus, the phase-plane approach is most useful in determining the response of a system to a step input. Since a step input does not always occur in practice, the application of the phase-plane technique is severely restricted when the response to other types of inputs is sought.

As an example, consider a second-order *linear* system whose characteristic equation is

$$\ddot{x} + 2\zeta\omega_n \dot{x} + \omega_n^2 x = 0 \qquad (10\text{-}9)$$

If the velocity $\dot{x}$ is treated as a new variable y, then, by eliminating the dependence of the above equation on time, there results

$$\frac{dy}{dx} + 2\zeta\omega_n + \omega_n^2 \frac{x}{y} = 0 \qquad (10\text{-}10)$$

Equation (10-10) is a first-order equation for y as a function of x and has a family of solutions depending on the initial values $y(0)$ and $x(0)$. Each solution is called a *phase trajectory,* and the totality of solutions is the phase-plane portrait of the system. The phase trajectories for Eq. (10-10) are shown in Fig. 10-15 for $\zeta = 0.5$. In this figure, if the initial conditions correspond to the point A_0, then the motion of the system is completely described by the trajectory $A_0A_1A_2A_3A_4A_5A_6$ with time increasing in the direction of the arrows.

In a more general case, it may be very difficult to solve the equation that describes the trajectories. A graphical procedure involving the determination of the *isoclines* (lines of equal slope) is then possible.[4] Referring to Eq. (10-8), the slope of the phase trajectories is found to be

$$\frac{dy}{dx} = -a(x,y) - b(x,y)\frac{x}{y} \qquad (10\text{-}11)$$

By setting the right side of this equation equal to a constant, a curve connecting points of equal slope is determined. The isoclines thus obtained are plotted in the phase plane, and the slopes of the various phase trajectories can be drawn directly on the isoclines. If a large number of isoclines are drawn, the phase trajectories can be accurately determined.

Once the phase portrait of a system has been constructed, the behavior of the system can be investigated. If the response of the system for a given set of initial conditions is sought, the corresponding phase trajectory determines the response. The variation of time t along the trajectory can be ascertained from the relation

$$t = \int \frac{1}{y}\, dx \qquad (10\text{-}12)$$

The nature of the stability of the system can be determined by an investigation of the *singular points* of the system. If the behavior

of a second-order system can be described by the two first-order equations

$$\dot{x} = P(x,y) \tag{10-13}$$

$$\dot{y} = Q(x,y) \tag{10-14}$$

the points at which $\dot{x} = 0$ and $\dot{y} = 0$ are called the *singular points* of the system and represent *equilibrium states* of the system. If the trajectories approach a singular point, the system is stable; whereas, if they diverge from the singular point, the system may be unstable. To investigate the nature of the equilibrium at a singular point, a Taylor series expansion of the functions $P(x,y)$ and $Q(x,y)$ is made about the point, and all but the first-order terms in expansion are neglected. Thus, the singular points are determined from the solutions of the equations

$$P(x,y) = 0 \tag{10-15}$$

$$Q(x,y) = 0 \tag{10-16}$$

The linearized forms of Eqs. **(10-13)** and **(10-14)** at the singular point $x = a$ and $y = b$ become

$$\dot{x} = a_1(x - a) + a_2(y - b) \tag{10-17}$$

$$\dot{y} = b_1(x - a) + b_2(y - b) \tag{10-18}$$

where a_1, a_2, b_1, and b_2 are coefficients of the expansion. Six types of singular points can occur:

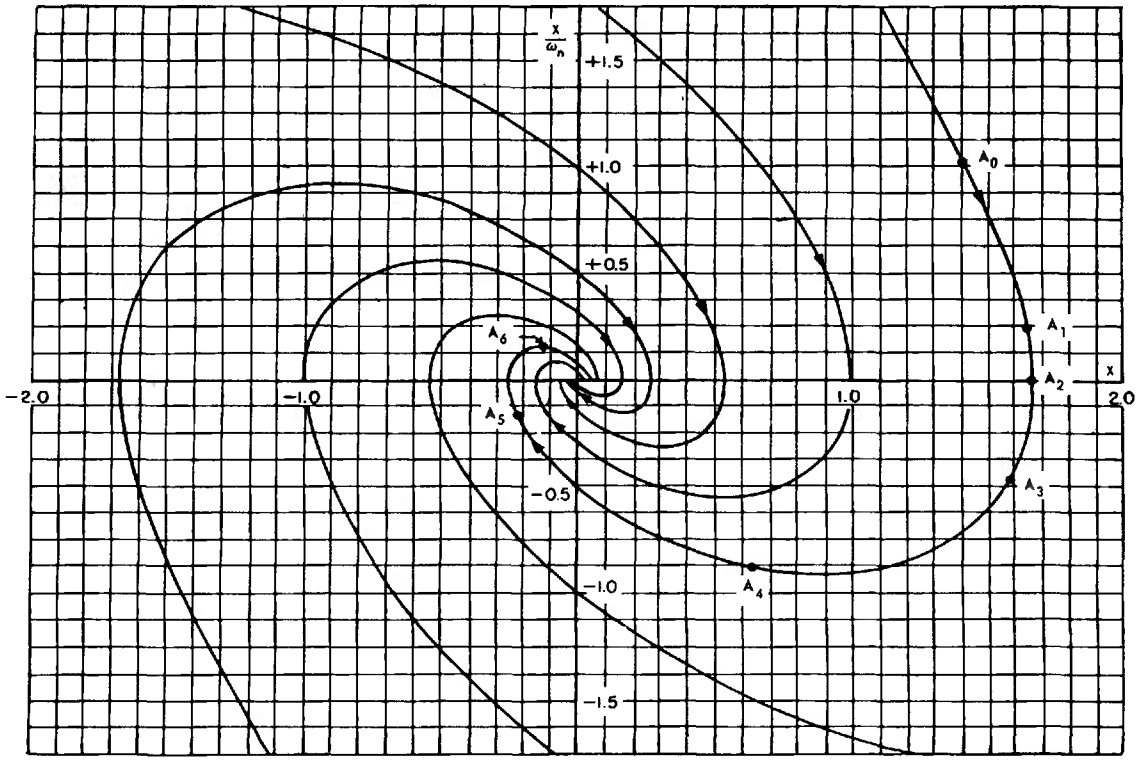


Fig. 10-15 Phase portrait of linear second-order system with $\zeta = 0.5$.

(a) *Stable node* (Fig. 10-16)

$$(a_1 + b_2) < 0 \qquad (10\text{-}19)$$

$$(a_1 + b_2)^2 > 4(a_1b_2 - a_2b_1) \qquad (10\text{-}20)$$

(b) *Unstable node* (Fig. 10-17)

$$(a_1 + b_2) > 0 \qquad (10\text{-}21)$$

$$(a_1 + b_2)^2 > 4(a_1b_2 - a_2b_1) \qquad (10\text{-}22)$$

(c) *Stable* focus (Fig. 10-18)

$$(a_1 + b_2) < 0 \qquad (10\text{-}23)$$

$$(a_1 + b_2)^2 < 4(a_1b_2 - a_2b_1) \qquad (10\text{-}24)$$

(d) *Unstable focus* (Fig. 10-19)

$$(a_1 + b_2) > 0 \qquad (10\text{-}25)$$

$$(a_1 + b_2)^2 < 4(a_1b_2 - a_2b_1) \qquad (10\text{-}26)$$

(e) *Center* (Fig. 10-20)

$$(a_1 + b_1) = 0 \qquad (10\text{-}27)$$

$$(a_1b_2 - a_2b_1) > 0 \qquad (10\text{-}28)$$

(f) *Saddle point* (Fig. 10-21)

$$(a_1 + b_1) = 0 \qquad (10\text{-}29)$$

$$(a_1b_2 - a_2b_1) < 0 \qquad (10\text{-}30)$$

The relations among the various singular points and the Taylor series coefficients given by Eqs. (10-19) through (10-30) are summarized in Fig. 10-22.

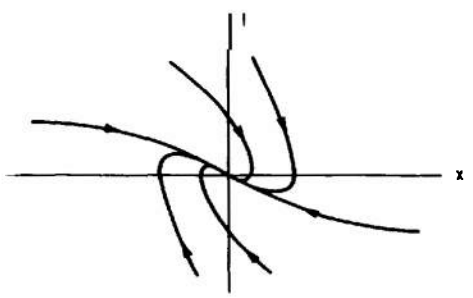


Fig. 10-16 Portrait in the vicinity of a stable node.

By permission from *Automatic Feedback Control System Synthesis*, by J. C. Truxal. Copyright. 1955, McGraw-Hill Book Company, Inc.

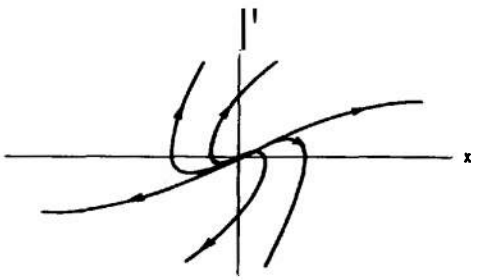


Fig. 10-17 Portrait in the vicinity of an unstable node.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

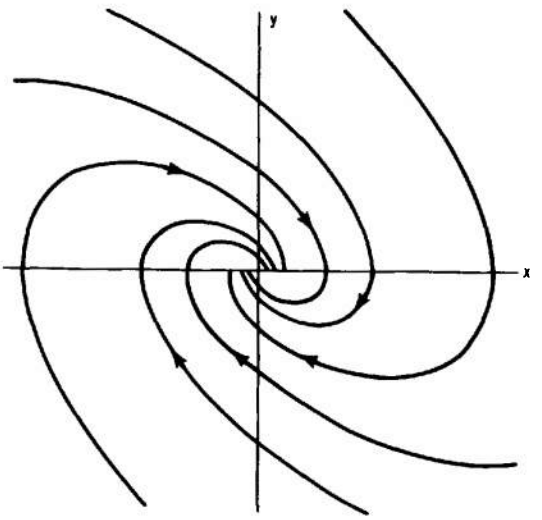


Fig. 70-78 Portrait in the vicinity of a stable focus.

By permission from *Automatic Feedback Control System Synthesis*. by J. G. Truxal, Copyright. 1955. McGraw-Hill Book Company, Inc.

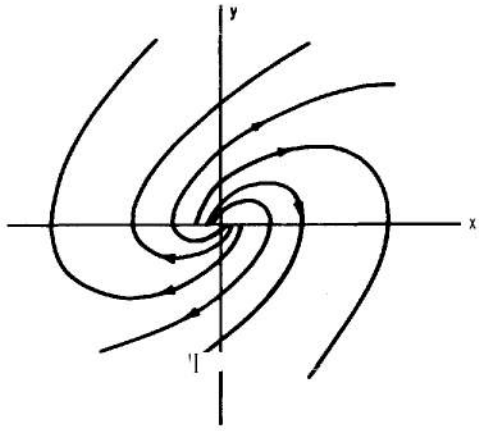


Fig. 10-19 Portrait in the vicinity of an unstable focus.

By permission from *Automatic Feedback Control System Synthesis.* by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

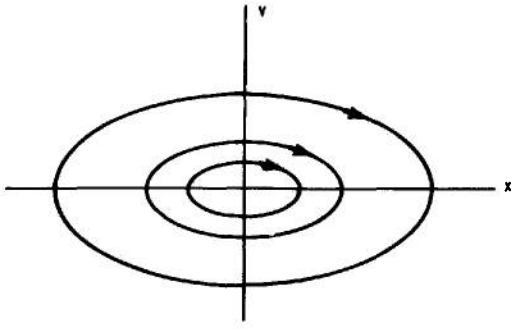


Fig. *10-20* Portrait in the vicinity of a center.

By permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal. Copyright, 1956. McGraw-Hill Book Company, Jnc.

In the case of feedback control systems, the problem is simplified because **Eq. (10-13)** can be replaced by

$$x = y \tag{10-31}$$

and the Taylor series coefficients a_1 and a_2 become

$$a_1 = 0 \tag{10-32}$$

$$a_2 = 1 \tag{10-33}$$

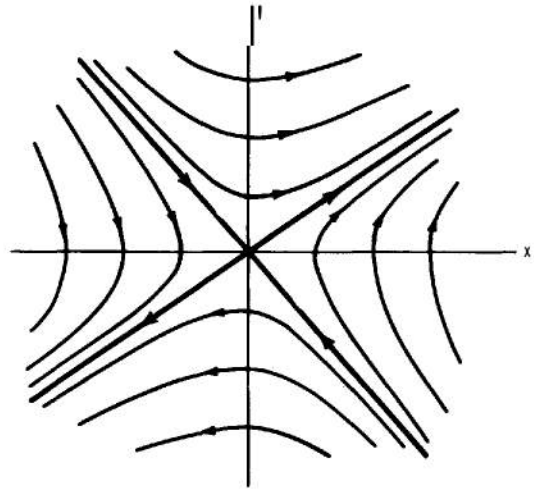


Fig. 10-21 Portrait in the neighborhood of a saddle point.

Adapted by permission from *Automatic Feedback Control System Syntheeis,* by J. G. Truxal, Copyrighty, 1955. McGraw-Hill Book Company, Inc.

In addition to the determination of the singular points, a complete description of the stability of a system in the phase plane requires a determination of the *limit cycles* of the system. *A limit cycle* is an isolated closed path in the phase portrait which corresponds to a system oscillation of fixed amplitude and period. A limit cycle is stable or unstable depending upon whether the paths in the neighborhood converge toward the limit cycle or diverge away from it. Thus, there arise two general types of self-excitation of nonlinear systems. *Soft excitation* occurs when a limit cycle encloses an unstable singular point (Fig. **10-23)** ; *hard excitation* **occurs** when a limit cycle encloses a stable limit cycle **or** a stable singular point (Fig. **10-24).** There is no definite method available for determining the limit cycles of a system or even if **a** limit cycle exists. The only approach is to determine the convergent and divergent properties of the phase trajectories. Thus, if all trajectories are converging outside a circle C_1 (centered at the origin) and diverging inside a smaller circle C_2 (centered at the origin), then a stable limit cycle must exist between

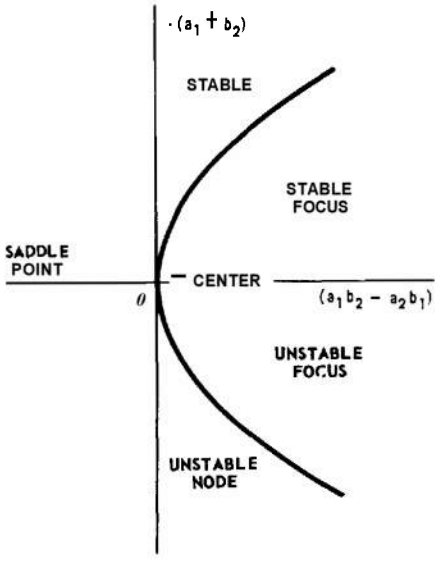


Fig. 70-22 Types of singularities.

By permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

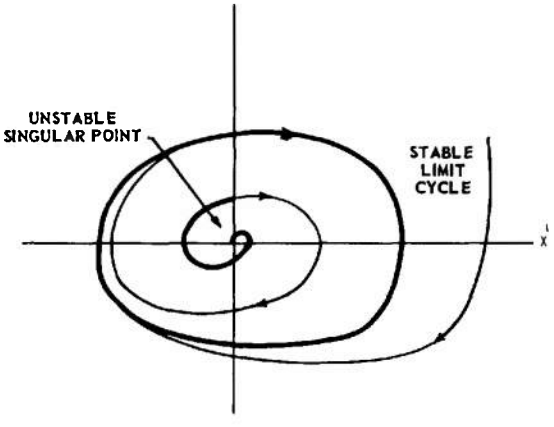


Fig. 10-23 Portrait with soft self-excitation.

By permission from *Automatic Feedback Control System Synthesis.* by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

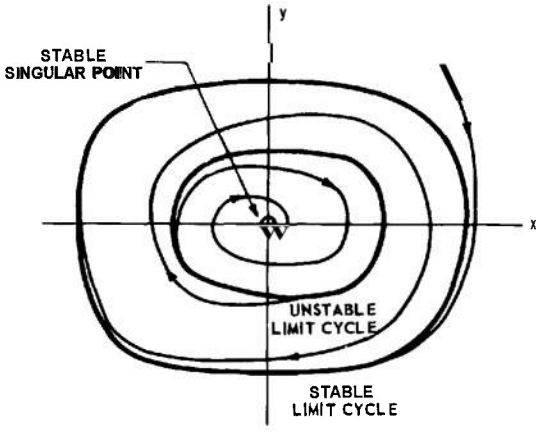


Fig. 10-24 Portrait with hard self-excitation.

By permission from *Automatic Feedback Control System Synthesis,* by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

the two circles. In particular, an examination of the time rate of change of the distance r from the origin for small and large values of x and y can determine the divergent or convergent properties of the phase trajectories. Several other conditions for the existence of limit cycles have been determined.[3] Some of these conditions are the following:

(a) No limit cycle exists in any region within which

$$\frac{\delta P}{\delta x} + \frac{\delta Q}{\delta y}$$

does not change sign.

(b) Within any limit cycle the number of nodes, foci, and centers must exceed the number of saddle points by one.

(c) If a trajectory stays inside a finite region and does not approach a singular point, then the trajectory must be a limit cycle or approach a limit cycle asymptotically.

Knowing the trajectories, the singular points, and the limit cycles of a system, the behavior of the system is completely determined when the initial conditions are specified. The determination of the trajectories and the singular points is a straightforward

procedure; the determination of the limit cycles, however, is more difficult. The complete phase portrait can then be used to determine the nature of system stability, and the response of the system (if stable) is readily ascertained.

10-4 LIMITATIONS, COMPENSATION, AND OTHER METHODS

As discussed previously, the describing-function procedure is primarily effective in determining the existence of limit cycles and predicting the amplitudes and frequencies associated with stable limit cycles. To a lesser degree the describing-function procedure can be used to estimate qualitatively the degree of stability of stable nonlinear systems. In addition, the describing-function method can be used to determine the frequency response of nonlinear systems; therefore, it is helpful in explaining anomalous experimental results.

The phase-plane procedure is useful in determining the *exact* nature of the stability of nonlinear systems in situations where the describing function method is inapplicable. In addition, the *time response* of a nonlinear system can be determined expeditiously through the use of the phase-plane so that a more quantitative estimate of the degree of stability of a system can be obtained.

Unfortunately, neither the phase plane nor the describing function can be used to determine the response of a nonlinear system to inputs other than simple steps or sinusoids. Since these elementary inputs rarely occur in practice, the utility of the two methods is severely restricted.

When the input to a system is arbitrarily defined, it is necessary to use either numerical computation[54,55,56] or, more conveniently, analog or digital computers. The analog computer is an especially powerful aid in the study of nonlinear systems.

Some specific remarks are in order regarding the stabilization and compensation of nonlinear systems. If the describing-function method is applicable, stabilization can be accomplished by reshaping the response $G(j\omega)$ of the linear element with conventional linear functions to eliminate intersections between the describing function and $G(j\omega)$. A nonlinear compensation function may be added to reshape the original describing function. If the added nonlinearity is separated from the original one by a low-pass filter, the describing functions of the two nonlinearities can be multiplied directly to obtain the composite describing function of the nonlinearly compensated system. If the compensating nonlinearity immediately precedes or follows the original nonlinearity with no separation by filtering action, a new describing function must be determined by combining the input-output characteristics of the two nonlinearities. In the latter case, the effect of the added nonlinearity on the original describing-function locus is much more difficult to visualize.

A great deal of effort has been devoted to the study of "optimum" nonlinear systems. The basic assumption in these studies[5,11,23,24,25,27,30,34,41,44,46,48] is that a system having a transient response (to a *step input)* that settles in a minimum period of time and has a minimum overshoot is an "optimum" system. The limitation of such "optimization" methods is due primarily to the fact that a nonlinear system will behave differently for different inputs. As a result, a system that has been "optimized" for a given step input may behave poorly for other step inputs of different magnitude, and it is likely that it will not behave in an optimum manner in response to other types of inputs.

In conclusion, it should be said that the problems of stabilization, compensation, and optimization of nonlinear systems have, as yet, not been adequately treated.

BIBLIOGRAPHY

1 H. T. Marcy, M. Yachter, and J. Zauderer, "Instrument Inaccuracies in Feedback Control Systems with Particular Reference to Backlash", *Trans. AIEE,* Vol. 68, Part I, pp. #778-788, 1949.

2 H. K. Weiss, "Analysis of Relay Servomechanisms", *J. of Aero. Sci.,* Vol. 13, No. 7, pp. #364-376, July, 1946.

3 N. Minorsky, *Introduction to Nonlinear Mechanics,* J. W. Edwards Bros., Ann Arbor, Mich., 1947.

4 A. A. Andronow and C. E. Chaikin, *Theory of Oscillations,* Princeton University Press, Princeton, N. J., 1949.

5 A. M. Hopkin, "A Phase-Plane Approach to the Compensation of Saturating Servomechanisms", *Trans. AIEE,* Vol. 70, Part I, pp. #631-639, 1951.

6 D. A. Kahn, "An Analysis of Relay Servomechanisms", *Trans. AIEE,* Vol. 68, Part 11, pp. #1079-1088, 1949.

7 R. J. Kochenburger, "A Frequency-Response Method for Analyzing and Synthesizing Contactor Servomechanisms", *Trans. AIEE,* Vol. 69, Part I, pp. #270-284, 1950.

8 A. Tustin, "The Effects of Backlash and of Speed-Dependent Friction on the Stability of Closed-Cycle Control Systems", *J. Inst. Elec. Engrs.* (London), Vol. 94, Part IIA, No. 1, pp. #143-151, May, 1947.

9 G. D. McCann, F. C. Lindvall, and C. H. Wilts, "The Effect of Coulomb Friction on the Performance of Servomechanisms", *Trans. AIEE,* Vol. 67, Part I, pp. #540-546, 1948.

10 C. Leondes and M. Rubinoff, "DINA, A Digital Analyzer for Laplace, Poisson, Diffusion, and Wave Equations", *Trans. AIEE,* Vol. 71, pp. #303-309, 1952.

11 J. B. Lewis, "The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism", *Trans. AIEE,* Vol. 71, Part 11, pp. #449-453, 1952.

12 N. Kryloff and No. Bogoliuboff, *Introduction to Nonlinear Mechanics,* translated by S. Lefschetz, Princeton University Press, Princeton, N. J., 1943.

13 B. V. Bulgakov, "Periodic Processes in Free Pseudo-Linear Oscillatory Systems", *J. Franklin Inst.,* Vol. 235, pp. #591-616, June, 1943.

14 J. C. Lozier, "Carrier-Controlled Relay Servos", *Electrical Engineering,* Vol. 69, No. 12, pp. #1052-1056, December, 1950.

15 E. C. Johnson, "Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements", *Trans. AIEE,* Vol. 71, Part 11, pp. #169-181, 1952.

16 E. S. Sherrard, "Stabilization of a Servomechanism Subject to Large Amplitude Oscillation", *Trans. AIEE,* Vol. 71, Part 11, pp. #312-324, 1952.

17 E. Levinson, "Some Saturation Phenomena in Servomechanisms with Emphasis on the Tachometer Stabilized System", *Trans. AIEE,* Vol. 72, Part 11, pp. #1-9, 1953.

18 V. B. Haas, Jr., "Coulomb Friction in Feedback Control Systems", *Trans. AIEE,* Vol. 72, Part 11, pp. #119-126, 1953.

19 R. J. Kochenburger, "Limiting in Feedback Control Systems", *Trans. AIEE,* Vol. 72, Part 11, pp. #180-194, 1953.

20 W. E. Scott, "An Introduction to the Analysis of Nonlinear Closed-Cycle Control Systems", in *Automatic and Manual Control,* edited by A. Tustin, pp. #249-261, Academic Press, Inc., New York, N.Y., 1952.

21 R. L. Cosgriff, "Open-Loop Frequency-Response Method for Nonlinear Servomechanisms", *Trans. AIEE,* Vol. 72, Part 11, pp. #222-225, 1953.

22 H. D. Greif, "Describing-Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities", *Trans. AIEE,* Vol. 72, Part 11, pp. #243-248, 1953.

23 R. S. Neiswander and R. H. MacNeal, "Optimization of Nonlinear Control Systems by Means of Nonlinear Feedbacks", *Trans. AIEE,* Vol. 72, Part II, pp. #262-272, 1953.

24 D. McDonald, "Nonlinear Techniques for Improving Servo Performance", in Proc. *Natl. Electronics Conf.,* Vol. 6, pp. #400-421, Chicago, Ill., 1950.

25 D. McDonald, "Multiple Mode Operation of Servomechanisms", *Rev. Sci. Instr.,* Vol. 23, pp. #22-30, January, 1952.

26 S. Lefschetz, *Contributions to the Theory of Nonlinear Oscillations,* Princeton University Press, Princeton, N. J., 1950.

27 L. F. Kazda, "Errors in Relay Servo Systems", *Trans. AIEE,* Vol. 72, Part 11, pp. #323-328, 1953.

28 T. A. Rogers and W. C. Hurty, "Relay Servomechanisms: The Shunt-Motor Servo with Inertia Load", *Trans. ASME,* Vol. 72, pp. #1163-1172, November, 1950.

29 A. M. Uttley and P. H. Hammond, "The Stabilization of On-Off Controlled Servomechanisms", in *Automatic and Manual Control,* edited by A. Tustin, pp. #285-307, Academic Press, New York, N.Y., 1952.

30 T. M. Stout, "Effects of Friction in an Optimum Relay Servomechanism", *Trans. AIEE,* Vol. 72, Part 11, pp. #329-336, 1953.

31 N. B. Nichols, "Backlash in a Velocity Lag Servomechanism", *Trans. AIEE,* Vol. 72, Part 11, pp. #462-467, 1953.

32 J. R. Burnett and P. E. Kendall, "Linear Compensation of Saturating Servomechanisms", *Trans. AIEE,* Vol. 73, Part 11, pp. #6-10, 1954.

33 C. K. Chow, "Contactor Servomechanisms Employing Sampled Data", *Trans. AIEE,* Vol. 73, Part 11, pp. #51-64, 1954.

34 I. Bogner and L. F. Kazda, "An Investigation of the Switching Criteria for Higher Order Contactor Servomechanisms", *Trans. AIEE,* Vol. 73, Part II, pp. #118-127, 1954.

35 L. T. Prince, Jr., "A Generalized Method for Determining the Closed-Loop Frequency Response of Nonlinear Systems", *Trans. AIEE,* Vol. 73, Part 11, pp. #217-224, 1954.

36 D. K. Gehmlich and M. E. Van Valkenberg, "Measurement of Some Nonlinearities in Servomechanisms", *Trans. AIEE,* Vol. 73, Part 11, pp. #232-235, 1954.

37 J. G. L. Michel and A. Porter, "The Effect of Friction on the Behavior of Servomechanisms at Creep Speeds", Proc. *Inst. Elec. Engrs.* (London), Vol. 98, Part 11, pp. #297-311, 1951.

38 H. Lauer, "Operating Modes of a Servomechanism with Nonlinear Friction", *J. Franklin Inst.,* Vol. 255, pp. #497-511, 1953.

39 Y. H. Ku, "Nonlinear Analysis of Electro-Mechanical Problems", *J. Franklin Inst.,* Vol. 255, pp. #9-31, 1953.

40 Y. H. Ku, "A Method for Solving Third and Higher Order Nonlinear Differential Equations", *J. Franklin Inst.,* Vol. 256, pp. #229-244, 1953.

41 R. E. Kalman, "Phase-Plane Analysis of Automatic Control Systems with Nonlinear Gain Elements", *Trans. AIEE,* Vol. 73, Part 11, pp. #383-390, 1954.

42 L. M. Vallese, "Analysis of Backlash in Feedback Control Systems with One Degree of Freedom", *Trans. AIEE,* Vol. 74, Part 11, pp. #I-4, 1955.

43 K. C. Matthews and R. C. Boe, "The Application of Nonlinear Techniques to Servomechanisms", *Proc. Natl. Electronics Conf.,* Vol. 8, pp. #IO-21, Chicago, Ill., 1952.

44 J. E. Hart, "An Analytical Method for the Design of Relay Servomechanisms", *Trans. AIEE,* Vol. 74, Part 11, pp. #83-90, 1955.

45 M. V. Matthews, "A Method for Evaluating Nonlinear Servomechanisms", *Trans. AIEE,* Vol. 74, Part 11, pp. #114-123, 1955.

46 S. S. L. Chang, "Optimum Switching Criteria for Higher Order Contactor Servo with Interrupted Circuits", *Trans. AIEE,* Vol. 74, Part 11, pp. #273-276, 1955.

47 Y. H. Ku, "Analysis of Nonlinear Systems with More than One Degree of Freedom by Means of Space Trajectories", *J. Franklin Inst.,* Vol. 259, pp. #115-131, February, 1955.

48 R. E. Kalman, "Analysis and Design Principles of Second and Higher Order Saturating Servomechanisms", *Trans. AIEE,* Vol. 74, Part 11, pp. #294-310, 1955.

49 R. C. Booton, Jr., "The Analysis of Nonlinear Control Systems with Random Inputs", *Proc. of the Symposium of Nonlinear Circuits Analysis,* pp. #369-391, Brooklyn Polytechnic Inst., Brooklyn, N. Y., April, 1953.

50 K. Chen, "Quasi-Linearization Techniques for Transient Study of Nonlinear Feedback Control Systems", *Trans. AIEE,* Vol. 74, Part 11, pp. #354-365, 1955.

51 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design,* Vol. 11, pp. #241-364, John Wiley & Sons, Inc., New York, N.Y., 1955.

52 J. G. Truxal, *Automatic Feedback Control System Synthesis,* pp. #559-663, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

53 I. Flugge-Lotz, *Discontinuous Automatic Control,* Princeton University Press, Princeton, N. J., 1953.

54 A. S. Householder, *Principles of Numerical Analysis,* McGraw-Hill Book Company, Inc., New York, N.Y., 1953.

55 F. B. Hildebrand, *Introduction to Numerical Analysis,* McGraw-Hill Book Company, Inc., New York, N.Y., 1956.

56 W. E. Milne, *Numerical Solutions of Differential Equations,* John Wiley & Sons, Inc., New York, N.Y., 1953.

☆ U.S. GOVERNMENT PRINTING OFFICE : 1968 O—300-512 (9976-A)

ENGINEERING DESIGN HANDBOOK SERIES

Listed below are the Handbooks which have been published or submitted for publication. Handbooks with publication dates prior to 1 August 1962 were published as 20-series Ordnance Corps pamphlets. AMC Circular 310-38, 19 July 1963, redesignated those publications as 706-series AMC pamphlets (i.e., ORDP 20-138 was redesignated AMCP 706-138). All new, reprinted, or revised Handbooks are being published as 706-series AMC pamphlets.

General and Miscellaneous Subjects

Number	Title
106	Elements of Armament Engineering, Part One, Sources of Energy
107	Elements of Armament Engineering, Part Two, Ballistics
108	Elements of Armament Engineering, Part Three, Weapon Systems and Components
110	Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data
111	Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data
112	Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments
113	Experimental Statistics, Section 4, Special Topics
114	Experimental Statistics, Section 5, Tables
121	Packaging and Pack Engineering
134	Maintenance Engineering Guide for Ordnance Design
135	Inventions, Patents, and Related Matters (Revised)
136	Servomechanisms, Section 1, Theory
137	Servomechanisms, Section 2, Measurement and Signal Converters
138	Servomechanisms, Section 3, Amplification
139	Servomechanisms, Section 4, Power Elements and System Design
170(C)	Armor and Its Application to Vehicles (U)
250	Guns--General (Guns Series)
252	Gun Tubes (Guns Series)
270	Propellant Actuated Devices
290(C)	Warheads--General (U)
331	Compensating Elements (Fire Control Series)
355	The Automotive Assembly (Automotive Series)

Ammunition and Explosives Series

175	Solid Propellants, Part One
176(C)	Solid Propellants, Part Two (U)
177	Properties of Explosives of Military Interest, Section 1
* 178(C)	Properties of Explosives of Military Interest, Section 2 (U)
210	Fuzes, General and Mechanical
211(C)	Fuzes, Proximity, Electrical, Part One (U)
212(S)	Fuzes, Proximity, Electrical, Part Two (U)
213(S)	Fuzes, Proximity, Electrical, Part Three (U)
214(S)	Fuzes, Proximity, Electrical, Part Four (U)
215(C)	Fuzes, Proximity, Electrical, Part Five (U)
244	Section 1, Artillery Ammunition--General, with Table of Contents, Glossary and Index for Series
245(C)	Section 2, Design for Terminal Effects (U)
246	Section 3, Design for Control of Flight Characteristics
247	Section 4, Design for Projection
248	Section 5, Inspection Aspects of Artillery Ammunition Design
249	Section 6, Manufacture of Metallic Components of Artillery Ammunition
*179	Explosive Trains

Ballistic Missile Series

Number	Title
281(S-RD)	Weapon System Effectiveness (U)
*282	Propulsion and Propellants
284(C)	Trajectories (U)
286	Structures
*283	Aerodynamics

Ballistics Series

140	Trajectories, Differential Effects, and Data for Projectiles
150	Interior Ballistics of Guns
160(S)	Elements of Terminal Ballistics, Part One, Introduction, Kill Mechanisms, and Vulnerability (U)
161(S)	Elements of Terminal Ballistics, Part Two, Collection and Analysis of Data Concerning Targets (U)
162(S-RD)	Elements of Terminal Ballistics, Part Three, Application to Missile and Space Targets (U)

Carriages and Mounts Series

340	Carriages and Mounts--General
341	Cradles
342	Recoil Systems
343	Top Carriages
344	Bottom Carriages
345	Equilibrators
346	Elevating Mechanisms
347	Traversing Mechanisms

Materials Handbooks

301	Aluminum and Aluminum Alloys
302	Copper and Copper Alloys
303	Magnesium and Magnesium Alloys
305	Titanium and Titanium Alloys
308	Glass
309	Plastics
310	Rubber and Rubber-Like Materials

Military Pyrotechnics Series

186	Part Two, Safety, Procedures and Glossary
187	Part Three, Properties of Materials Used in Pyrotechnic Compositions

Surface-to-Air Missile Series

291	Part One, System Integration
292	Part Two, Weapon Control
293	Part Three, Computers
294(S)	Part Four, Missile Armament (U)
295(S)	Part Five, Countermeasures (U)
296	Part Six, Structures and Power Sources
297(S)	Part Seven, Sample Problem (U)